



ANALYSIS AND FORECASTING OF MULTI-REGION RECYCLABLE WASTE DATASETS WITH SUPERVISED MACHINE LEARNING MODELS

GÖZDE AKA BAL

Thesis for Master's Program in Industrial Engineering

Graduate School

Izmir University of Economics

İzmir

2024

ANALYSIS AND FORECASTING OF MULTI-REGION RECYCLABLE WASTE DATASETS WITH SUPERVISED MACHINE LEARNING MODELS

GÖZDE AKA BAL

THESIS ADVISOR: ASST. PROF. DR. KAMİL ERKAN KABAK

Master's Exam Jury Members

Assoc. Prof. Dr. Önder BULUT

Asst. Prof. Dr. Alper DEMİR

Asst. Prof. Dr. Kamil Erkan KABAK

A Master's Thesis

Submitted to

the Graduate School of Izmir University of Economics

the Department of Industrial Engineering

İzmir

2024

ETHICAL DECLARATION

I hereby declare that I am the sole author of this thesis and that I have conducted my work in accordance with academic rules and ethical behaviour at every stage from the planning of the thesis to its defence. I confirm that I have cited all ideas, information and findings that are not specific to my study, as required by the code of ethical behaviour, and that all statements not cited are my own.

Name, Surname:

Gözde AKBA

Date:

19.04.2024

ABSTRACT

ANALYSIS AND FORECASTING OF MULTI-REGION RECYCLABLE WASTE DATASETS WITH SUPERVISED MACHINE LEARNING MODELS

Aka Bal, Gözde

Master's Program in Industrial Engineering

Advisor: Asst. Prof. Dr. Kamil Erkan KABAK

April, 2024

To have sustainable cities, it is quite crucial to forecast the recyclable waste amounts at different points of regions so that the return of some portions of such amounts would allow to minimize the raw material usage at manufacturing facilities and increase profit via recycling. For these reasons, the objective of this study is to analyze the amount of recycled waste collected at different time intervals from different regions, after that it aims to forecast the waste amounts through different supervised machine-learning approaches. These forecasts will benefit organizations to better shape their recycling strategies and minimize costs through optimal recycling collection and investment policies. Due to the uncertainties and variabilities within the collected recycled waste datasets, this thesis first applies traditional forecasting

methods with time series analyses. Then, further analyses are performed through supervised machine learning approaches such as ARIMA, K-NN, and XGBoost apart from statistical analyses that cover descriptive and correlation analyses. Based on the analyses conducted by traditional forecasting methods, the moving average method provides the best forecasting results in all regions with MSE value of 97 among other methods for paper waste and 0.13 for plastic waste. Concerning the analyses by supervised machine learning methods, the ARIMA approach shows better forecasting results in the majority of the regions, and it has with MSE value of 105 for the best-performing region. Finally, the study proposes an automated forecasting framework after iteratively combining the forecasting algorithms in a pool and selecting the best forecasts according to predefined performance measures.

Keywords: recyclable waste, forecasting, machine learning approaches, statistical analysis.

ÖZET

DENETİMLİ MAKİNE ÖĞRENME MODELLERİ İLE ÇOK BÖLGELİ GERİ DÖNÜŞTÜRÜLEBİLİR ATIK VERİ SETLERİNİN ANALİZİ VE ÖNGÖRÜLMESİ

Aka Bal, Gözde

Endüstri Mühendisliği Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Kamil Erkan Kabak

Nisan, 2024

Sürdürülebilir şehirler için, bölgelerin farklı atık miktarlarını tahmin etmek oldukça önemlidir; böylelikle bu miktarların bir kısmının geri dönüşü, üretim tesislerinde hammadde kullanımını enküçükleyebilecektir ve geridönüşümle karlılığı arttırabilecektir. Bu nedenlerle, bu çalışmanın amacı, farklı zaman aralıklarında ve farklı bölgelerde toplanan geri dönüşüm atık miktarlarını analiz etmektir, sonrasında farklı gözetimli makine öğrenimi yaklaşımları ile atık miktarlarını tahmin etmeyi amaçlamaktadır. Bu tahminler kuruluşların geri dönüşüm stratejilerini daha iyi tahminlerle daha iyi şekillendirmelerine ve en iyi geri dönüşüm toplama ve yatırım politikaları ile maliyetlerinin minimize etmeyi sağlayacaktır. Farklı noktalardan elde edilen geri dönüşüm atık veri setlerindeki belirsizliklerden ve değişkenliklerden dolayı, bu tez

alışması öncelikle geleneksel tahminleme yöntemlerini zaman serisi analizleri ile uygulamaktadır. Sonra, tanımlayıcı ve korrelasyon analizlerini içeren istatistiksel analizlerin dışında ARIMA, K-NN ve XGBoost gibi seçilmiş gözetimli makine öğrenmesi yaklaşımları ile ilave analizleri yapmaktadır. Geleneksel tahmin yöntemleriyle yapılan analizlere göre hareketli ortalama yöntemi, kağıt atıklarında diğer yöntemler arasında 97, plastik atıklarda ise 0,13 MSE değeriyle tüm bölgelerde en iyi tahmin sonuçlarını sağlamaktadır. Denetimli makine öğrenimi yöntemleri analizlerine göre, ARIMA yöntemi bölgelerin çoğunda daha iyi tahminleme sonuçlarını göstermektedir ve en iyi performans gösteren bölge 105 MSE değerine sahiptir. Son olarak, çalışma tahminleme algoritmalarını bir havuzda toplayarak ve en iyi tahminleri tanımlanan tahminleme performansı ölçütlerine göre seçerek otomatik bir tahminleme çerçevesi oluşturmayı amaçlamaktadır.

Anahtar Kelimeler: geri-dönüştürülebilir atık, tahminleme, makine öğrenimi yaklaşımları, istatistiksel analiz.

This thesis work is dedicated to my family.



ACKNOWLEDGEMENT

Firstly, I would like to thank to my advisor, Asst. Prof. Dr. Kamil Erkan Kabak for his guidance, encouragement and patience during the course of this thesis.

I want to thank Assoc. Prof. Dr. Önder Bulut and Asst. Prof. Dr. Alper Demir, who agreed to take part in the thesis jury, for their suggestions and comments at every stage.

I would like to thank Ilker Teknik Metal Makine Ltd. Co., located in Izmir, for providing the dataset for this thesis.

I am very thankful to my father Dr. Murat Tamer Aka and my mother Filiz Aka for their endless trust and belief that I can complete this process successfully. Also, I am very grateful to my dear sister Dr. Hande Aka Uymaz, who shared her excellent knowledge and always supported me.

Lastly, I would like to thank to my precious husband Sami Bal for his endless love, support and motivation.

TABLE OF CONTENTS

ABSTRACT.....	iv
ÖZET	vi
DEDICATION	viii
ACKNOWLEDGEMENT	ix
TABLE OF CONTENTS.....	x
LIST OF TABLES	xiii
LIST OF FIGURES	xxiii
CHAPTER 1: INTRODUCTION.....	1
1.1. <i>Aim and Scope of the Thesis</i>	1
1.2. <i>Problem Definition</i>	4
1.3. <i>Research Methodology</i>	4
1.4. <i>Summary of Chapters in the Study</i>	7
CHAPTER 2: LITERATURE RESEARCH.....	8
CHAPTER 3: CASE STUDY.....	14
CHAPTER 4: METHODOLOGY	16
4.1. <i>Data Collection</i>	17
4.2. <i>Dataset Analysis of Pre-Processing in Real Dataset</i>	18
4.3. <i>Analyses of Descriptive Statistics of Waste Datasets</i>	22
4.4. <i>The Correlation Chart</i>	29
4.5. <i>Concluding Remarks</i>	30
CHAPTER 5: APPLICATION OF TRADITIONAL FORECASTING METHODS	32
5.1. <i>Model Performance Evaluation</i>	32
5.2. <i>Moving Average Method (MA)</i>	33
5.3. <i>Weighted Moving Average Method (WMA)</i>	36
5.4. <i>Exponential Smoothing Method (ES)</i>	40
5.5. <i>Holt-Winters Method (HW)</i>	42
5.6. <i>Comparison of Model Results Predicted by Traditional Forecast Models</i>	47
CHAPTER 6: FORECASTING WITH MACHINE LEARNING	49

6.1. Supervised Learning and Selected Methods for Analysis	49
6.1.1. Extreme Gradient Boosting (XGBoost).....	50
6.1.2. K-Nearest Neighbors (K-NN).....	51
6.1.3. Autoregressive Integrated Moving Average (ARIMA)	52
6.2. Validation of Models	53
6.3. Comparison of Model Results Forecasted by Machine Learning	54
6.3.1. Extreme Gradient Boosting (XGBoost) Results	54
6.3.2. K-Nearest Neighbors (K-NN) Results.....	61
6.3.3. Autoregressive Integrated Moving Average (ARIMA) Results	68
6.3.4. Interpreting the Results of Supervised Machine Learning Models	74
6.4. Comparison of Machine Learning and Traditional Methods.....	76
CHAPTER 7: AUTOMATION FRAMEWORK FOR FORECASTING WASTES	80
7.1. Analysis Steps.....	80
CHAPTER 8: CONCLUSION	87
REFERENCES	95
APPENDICES	95
Appendix A - Histogram and Scatter Line of Wastes in Regions	97
Appendix B - Moving Average Method for Wastes in Regions	105
Appendix C - Weighted Moving Average for Wastes in Regions	111
Appendix D - Exponential Smoothing Results for Wastes in Regions	117
Appendix E - Holt-Winters Results for Wastes in Regions	132
Appendix F - Graphs of Actual and Forecasted Values using XGBoost.....	137
Appendix G - Graphs of Actual and Forecasted Values using K-NN.....	143
Appendix H - Graphs of Actual and Forecasted Values using ARIMA	149

LIST OF TABLES

Table 1.	Summary Table of the Proposed Methods from Literature.....	12
Table 2.	Packaging Waste Cumulative Collected Data Related Production, Use and Recycle in Turkey Between 2014 to 2018.....	17
Table 3.	Total Amounts of Paper and Plastic Waste Collected in Each Region ..	19
Table 4.	Statistical Analysis of Paper Waste for Each Region	22
Table 5.	Statistical Analysis of Plastic Waste for Each Region.....	23
Table 6.	Fit Distributions and Parameters of Paper Waste for Each Region.....	26
Table 7.	Fit Distributions and Parameters of Plastic Waste for Each Region	27
Table 8.	Moving Average Method Forecast Results for Paper and Plastic Waste	36
Table 9.	Weighted Moving Average Method Results for Paper and Plastic Waste	39
Table 10.	Exponential Smoothing Method Results for Paper and Plastic Waste...	42
Table 11.	Holt-Winters Results of Paper and Plastic Waste for Each Region	47
Table 12.	Comparison of Consistencies of Methods for Paper Waste	48
Table 13.	Comparison of Consistencies of Methods for Plastic Waste.....	48
Table 14.	Train Set Results of Paper Waste using XGBoost	55
Table 15.	Average MSE Values Obtained From 5-Fold Cross-Validation for Paper Waste	56
Table 16.	Paper Waste Forecasting Results using XGBoost	58
Table 17.	Train Set Results for Plastic Waste using XGBoost	58
Table 18.	Average MSE Values Obtain From 5-Fold Cross Validation for Plastic Waste	59
Table 19.	Plastic Waste Forecasting Results using XGBoost.....	61
Table 20.	Train Set Results of Paper Waste using K-NN	62
Table 21.	Average MSE Values Obtain From 5-Fold Cross-Validation for Paper Waste.....	63
Table 22.	Paper Waste Forecasting Results Obtained from Testset using K-NN ..	65
Table 23.	Trainset Results of Plastic Waste using K-NN	65

Table 24. Average MSE Values Obtain From 5-Fold Cross-Validation for Plastic Waste	66
Table 25. Plastic Waste Forecasting Results Obtained from Testset using K-NN .	68
Table 26. Trainset Results of Paper Waste using ARIMA	69
Table 27. Paper Waste Forecasting Results Obtained from Testset using ARIMA	71
Table 28. Trainset Results for Plastic Waste using ARIMA	71
Table 29. Plastic Waste Forecasting Results Obtained from Testset using ARIMA	74
Table 30. Comparison of Consistencies of Machine Learning Models for Paper Waste.....	75
Table 31. Comparison of Consistencies of Machine Learning Models for Plastic Waste	76
Table 32. Comparison of All Methods for Paper Waste	78
Table 33. Comparison of All Methods for Plastic Waste	79

LIST OF FIGURES

Figure 1.	A Flow Chart for the Stages of Waste Data Used in This Study	15
Figure 2.	Flow Chart of Methodology	16
Figure 3.	Total Number of Packaging Waste at All Stages Between 2014-2018	18
Figure 4.	Total Amounts of Paper Waste by Regions	19
Figure 5.	Total Amounts of Plastic Waste by Regions	20
Figure 6.	Total Amounts of Waste Collected Monthly	20
Figure 7.	Distribution of Average Paper Waste Data by Days of the Week	21
Figure 8.	Distribution of Average Plastic Waste Data by Days of the Week	22
Figure 9.	Box Plot of Paper Waste by Regions	24
Figure 10.	Box Plot of Plastic Waste by Regions	25
Figure 11.	Histogram and Curve Fitting Line of Paper Waste in Regions	26
Figure 12.	Histogram and Curve Fitting Line of Plastic Waste in Regions	27
Figure 13.	Amount of Paper Waste in Multiple Regions	28
Figure 14.	Amount of Both Paper and Plastic Waste in Region A	29
Figure 15.	Correlation Matrix Between Regions	30
Figure 16.	Moving Average Method Using 3 & 6 Days Period in Region A	34
Figure 17.	Moving Average Method Using 3 & 6 Days Period in Region B.....	35
Figure 18.	Weighted Moving Average for Paper Waste in Region A	38
Figure 19.	Weighted Moving Average for Plastic Waste in Region A	39
Figure 20.	Exponential Smoothing Results for Paper Waste in Region A	41
Figure 21.	Exponential Smoothing Results for Plastic Waste in Region A	41
Figure 22.	Decomposition of Time Series Model for Paper Waste in Region A ...	44
Figure 23.	Holt-Winters Results for Paper Waste in Region A	45
Figure 24.	Decomposition of Time Series Model for Plastic Waste in Region A ..	46
Figure 25.	Holt-Winters Results for Plastic Waste in Region A	46
Figure 26.	Graphs of Actual and Forecasted Values in the Testset of Region H ...	56
Figure 27.	Graphs of Actual and Forecasted Values in the Testset of Region C....	57
Figure 28.	Graphs of Actual and Forecasted Values in the Testset of Region A ...	57

Figure 29. Graphs of Actual and Forecasted Values in the Testset of Region G ...	59
Figure 30. Graphs of Actual and Forecasted Values in the Test Set of Region J ...	60
Figure 31. Graphs of Actual and Forecasted Values in the Testset of Region B....	60
Figure 32. Graph of Forecasted Data Corresponding to Actual Data of Paper Waste in Region D.....	63
Figure 33. Graph of Forecasted Data Corresponding to Actual Data of Paper Waste in Region A.....	64
Figure 34. Graph of Forecasted Data Corresponding to Actual Data of Paper Waste in Region H.....	64
Figure 35. Graph of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region L	66
Figure 36. Graph of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region J.....	67
Figure 37. Graph of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region B	67
Figure 38. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region L	69
Figure 39. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region I	70
Figure 40. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region J	70
Figure 41. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region L	72
Figure 42. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region J	73
Figure 43. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region B	73
Figure 44. Paper Waste Forecasting Results of Each Region	75
Figure 45. Steps of Analysis of Automation Framework for Forecasting.....	82
Figure 46. Histogram and Curve Fitting Line of Paper Waste in Regions E-H.....	96
Figure 47. Histogram and Curve Fitting Line of Paper Waste in Regions I-L	96

Figure 48. Histogram and Curve Fitting Line of Plastic Waste in Regions E-H....	97
Figure 49. Histogram and Curve Fitting Line of Plastic Waste in Regions I-L.....	97
Figure 50. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region B.....	98
Figure 51. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region C.....	98
Figure 52. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region D.....	98
Figure 53. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region E.....	99
Figure 54. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region F.....	99
Figure 55. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region G.....	99
Figure 56. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region H.....	100
Figure 57. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region I.....	100
Figure 58. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region J.....	100
Figure 59. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region K.....	101
Figure 60. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region L.....	101
Figure 61. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region B.....	101
Figure 62. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region C.....	102
Figure 63. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region D.....	102

Figure 64. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region E	102
Figure 65. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region F	103
Figure 66. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region G.....	103
Figure 67. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region H.....	103
Figure 68. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region I.....	104
Figure 69. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region J.....	104
Figure 70. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region K.....	104
Figure 71. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region L.....	105
Figure 72. Weighted Moving Average for Paper Waste in Region B.....	106
Figure 73. Weighted Moving Average for Paper Waste in Region C.....	106
Figure 74. Weighted Moving Average for Paper Waste in Region D	106
Figure 75. Weighted Moving Average for Paper Waste in Region E.....	107
Figure 76. Weighted Moving Average for Paper Waste in Region F.....	107
Figure 77. Weighted Moving Average for Paper Waste in Region G	107
Figure 78. Weighted Moving Average for Paper Waste in Region H	107
Figure 79. Weighted Moving Average for Paper Waste in Region I.....	108
Figure 80. Weighted Moving Average for Paper Waste in Region J	108
Figure 81. Weighted Moving Average for Paper Waste in Region K	108
Figure 82. Weighted Moving Average for Paper Waste in Region L.....	108
Figure 83. Weighted Moving Average for Plastic Waste in Region B	109
Figure 84. Weighted Moving Average for Plastic Waste in Region C	109
Figure 85. Weighted Moving Average for Plastic Waste in Region D	109
Figure 86. Weighted Moving Average for Plastic Waste in Region E	109

Figure 87. Weighted Moving Average for Plastic Waste in Region F.....	110
Figure 88. Weighted Moving Average for Plastic Waste in Region G	110
Figure 89. Weighted Moving Average for Plastic Waste in Region H	110
Figure 90. Weighted Moving Average for Plastic Waste in Region I	110
Figure 91. Weighted Moving Average for Plastic Waste in Region J	111
Figure 92. Weighted Moving Average for Plastic Waste in Region K	111
Figure 93. Weighted Moving Average for Plastic Waste in Region L	111
Figure 94. Exponential Smoothing Results for Paper Waste in Region B	112
Figure 95. Exponential Smoothing Results for Paper Waste in Region C	112
Figure 96. Exponential Smoothing Results for Paper Waste in Region D	112
Figure 97. Exponential Smoothing Results for Paper Waste in Region E	113
Figure 98. Exponential Smoothing Results for Paper Waste in Region F.....	113
Figure 99. Exponential Smoothing Results for Paper Waste in Region G	113
Figure 100. Exponential Smoothing Results for Paper Waste in Region H	113
Figure 101. Exponential Smoothing Results for Paper Waste in Region I	114
Figure 102. Exponential Smoothing Results for Paper Waste in Region J	114
Figure 103. Exponential Smoothing Results for Paper Waste in Region K	114
Figure 104. Exponential Smoothing Results for Paper Waste in Region L	114
Figure 105. Exponential Smoothing Results for Plastic Waste in Region B	115
Figure 106. Exponential Smoothing Results for Plastic Waste in Region C	115
Figure 107. Exponential Smoothing Results for Plastic Waste in Region D.....	115
Figure 108. Exponential Smoothing Results for Plastic Waste in Region E	115
Figure 109. Exponential Smoothing Results for Plastic Waste in Region F	116
Figure 110. Exponential Smoothing Results for Plastic Waste in Region G.....	116
Figure 111. Exponential Smoothing Results for Plastic Waste in Region H.....	116
Figure 112. Exponential Smoothing Results for Plastic Waste in Region I	116
Figure 113. Exponential Smoothing Results for Plastic Waste in Region J	117
Figure 114. Exponential Smoothing Results for Plastic Waste in Region K.....	117
Figure 115. Exponential Smoothing Results for Plastic Waste in Region L	117
Figure 116. Decomposition of Time Series Model for Paper Waste in Region B ...	118
Figure 117. Holt-Winters Results for Paper Waste in Region B	118

Figure 118. Decomposition of Time Series Model for Paper Waste in Region C ...	118
Figure 119. Holt-Winters Results for Paper Waste in Region C	119
Figure 120. Decomposition of Time Series Model for Paper Waste in Region D ...	119
Figure 121. Holt-Winters Results for Paper Waste in Region D	119
Figure 122. Decomposition of Time Series Model for Paper Waste in Region E....	120
Figure 123. Holt-Winters Results for Paper Waste in Region E	120
Figure 124. Decomposition of Time Series Model for Paper Waste in Region F....	120
Figure 125. Holt-Winters Results for Paper Waste in Region F.....	121
Figure 126. Decomposition of Time Series Model for Paper Waste in Region G ...	121
Figure 127. Holt-Winters Results for Paper Waste in Region G	121
Figure 128. Decomposition of Time Series Model for Paper Waste in Region H ...	122
Figure 129. Holt-Winters Results for Paper Waste in Region H	122
Figure 130. Decomposition of Time Series Model for Paper Waste in Region I....	122
Figure 131. Holt-Winters Results for Paper Waste in Region I	123
Figure 132. Decomposition of Time Series Model for Paper Waste in Region J	123
Figure 133. Holt-Winters Results for Paper Waste in Region J	123
Figure 134. Decomposition of Time Series Model for Paper Waste in Region K ...	124
Figure 135. Holt-Winters Results for Paper Waste in Region K	124
Figure 136. Decomposition of Time Series Model for Paper Waste in Region L....	124
Figure 137. Holt-Winters Results for Paper Waste in Region L	125
Figure 138. Decomposition of Time Series Model for Plastic Waste in Region B ..	125
Figure 139. Holt-Winters Results for Plastic Waste in Region B	125
Figure 140. Decomposition of Time Series Model for Plastic Waste in Region C ..	126
Figure 141. Holt-Winters Results for Plastic Waste in Region C	126
Figure 142. Decomposition of Time Series Model for Plastic Waste in Region D ..	126
Figure 143. Holt-Winters Results for Plastic Waste in Region D.....	127
Figure 144. Decomposition of Time Series Model for Plastic Waste in Region E ..	127
Figure 145. Holt-Winters Results for Plastic Waste in Region E	127
Figure 146. Decomposition of Time Series Model for Plastic Waste in Region F ..	128
Figure 147. Holt-Winters Results for Plastic Waste in Region F	128
Figure 148. Decomposition of Time Series Model for Plastic Waste in Region G ..	128

Figure 149. Holt-Winters Results for Plastic Waste in Region G.....	129
Figure 150. Decomposition of Time Series Model for Plastic Waste in Region H ..	129
Figure 151. Holt-Winters Results for Plastic Waste in Region H.....	129
Figure 152. Decomposition of Time Series Model for Plastic Waste in Region I ...	130
Figure 153. Holt-Winters Results for Plastic Waste in Region I.....	130
Figure 154. Decomposition of Time Series Model for Plastic Waste in Region J ...	130
Figure 155. Holt-Winters Results for Plastic Waste in Region J.....	131
Figure 156. Decomposition of Time Series Model for Plastic Waste in Region K..	131
Figure 157. Holt-Winters Results for Plastic Waste in Region K.....	131
Figure 158. Decomposition of Time Series Model for Plastic Waste in Region L ..	132
Figure 159. Holt-Winters Results for Plastic Waste in Region L	132
Figure 160. Graphs of Actual and Forecasted Values for Paper Waste in Region B	133
Figure 161. Graphs of Actual and Forecasted Values for Paper Waste in Region D	133
Figure 162. Graphs of Actual and Forecasted Values for Paper Waste in Region E	133
Figure 163. Graphs of Actual and Forecasted Values for Paper Waste in Region F.	134
Figure 164. Graphs of Actual and Forecasted Values for Paper Waste in Region G	134
Figure 165. Graphs of Actual and Forecasted Values for Paper Waste in Region I .	134
Figure 166. Graphs of Actual and Forecasted Values for Paper Waste in Region J .	134
Figure 167. Graphs of Actual and Forecasted Values for Paper Waste in Region K	135
Figure 168. Graphs of Actual and Forecasted Values for Paper Waste in Region L	135
Figure 169. Graphs of Actual and Forecasted Values for Plastic Waste in Region A	135
Figure 170. Graphs of Actual and Forecasted Values for Plastic Waste in Region C	135
Figure 171. Graphs of Actual and Forecasted Values for Plastic Waste in Region D	136
Figure 172. Graphs of Actual and Forecasted Values for Plastic Waste in Region E	136
Figure 173. Graphs of Actual and Forecasted Values for Plastic Waste in Region F	136
Figure 174. Graphs of Actual and Forecasted Values for Plastic Waste in Region H	136
Figure 175. Graphs of Actual and Forecasted Values for Plastic Waste in Region I	137
Figure 176. Graphs of Actual and Forecasted Values for Plastic Waste in Region K	137
Figure 177. Graphs of Actual and Forecasted Values for Plastic Waste in Region L	137
Figure 178. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region B	138

Figure 179. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region C	138
Figure 180. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region E	138
Figure 181. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region F	139
Figure 182. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region G	139
Figure 183. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region I	139
Figure 184. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region J	140
Figure 185. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region K	140
Figure 186. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region L	140
Figure 187. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region A	141
Figure 188. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region C	141
Figure 189. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region D	141
Figure 190. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region E	142
Figure 191. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region F	142
Figure 192. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region G	142
Figure 193. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region H	143

Figure 194. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region I	143
Figure 195. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region K.....	143
Figure 196. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region A.....	144
Figure 197. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region B	144
Figure 198. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region C	144
Figure 199. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region D	145
Figure 200. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region E	145
Figure 201. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region F	145
Figure 202. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region G.....	146
Figure 203. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region H.....	146
Figure 204. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region K.....	146
Figure 205. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region A.....	147
Figure 206. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region C	147
Figure 207. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region D.....	147
Figure 208. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region E	148

Figure 209. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region F	148
Figure 210. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region G.....	148
Figure 211. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region H.....	149
Figure 212. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region I	149
Figure 213. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region K.....	149



CHAPTER 1: INTRODUCTION

In this chapter, Section 1.1 gives the aim and scope of this thesis study, then Section 1.2 presents the problem definition, this is followed by a brief description of the research methodologies in Section 1.3, and finally, Section 1.4 gives brief summaries of the remaining chapters.

1.1. Aim and Scope of the Thesis

The manufactured products have environmental effects at all stages. Improper disposal methods for waste are dangerous for human health and increase environmental pollution (Hossain et al., 2011). Increasing environmental problems cause societies to focus more seriously on these problems such as air pollution and water pollution. In addition, limited resources and rapid consumption due to population growth reveal the concept of sustainability. The number of studies on the reuse of wastes that consider them as sustainable resources is reported to be increasing apart from its importance (Solak and Pekküçükşen, 2018). With the recycling of waste, both waste disposal and raw material costs can be reduced. While minimizing the damage of waste to the environment, companies can make profits with a sustainability strategy. According to the study conducted in the United States, it was concluded that a 1 percent increase in waste recycling contributes 0.317 percent to economic growth (Razzaq et al., 2021). In Turkey, according to the data of the Ministry of Environment and Urbanization, 2 billion 188 million 739 thousand Turkish liras contributed to the economy with 1 million 836 thousand 279 tons of recycled packaging waste in 2020.

Waste forecasting has an important role in the field of waste management and can contribute to various decision-making processes (Birgen et al., 2021; Abdella Ahmed et al., 2022). Waste forecasting can help predict waste generation and a facility's waste stock. This information can be used in the facility's inventory management and production planning processes. These facilities may include a variety of facilities such as waste management facilities, recycling facilities, industrial facilities, manufacturing facilities and businesses in the service sector. For example,

waste forecasting could be crucial for determining a facility's waste storage capacity and keeping waste stock to a minimum. Forecasting waste amounts have a critical role in creating production planning schedules and using resources efficiently. To illustrate, waste forecasting predicts waste generation and can help adjust production schedules based on these waste amounts. This allows production processes to keep waste generation to a minimum and reduce waste management costs. Additionally, these forecasts can be used in waste transportation and logistics planning processes. Based on forecasted waste quantities, the resources and logistics operations required to collect, transport and dispose of waste can be planned more effectively (Seçkiner et al., 2021). Specifically, Vehicle Routing Problems (VRPs) may arise due to waste collection times, waste transportation from different locations, and special waste types each of which may have specific time and transportation mode constraints. Transportation costs related to waste characteristics could also be a concern. In addition, it can be used to increase the effectiveness of recycling and resource management programs. Accurate waste forecasting could have a major role for planning the capacity of recycling facilities, setting recycling targets and ensuring more efficient use of resources (Birgen et al., 2021). Waste forecasting can also be used in the processes of determining waste volume and composition. This information can be taken into account in the development of waste management strategies and policies and in the design of waste disposal facilities. For example, Birgen et al. (2021) work on waste management employing traditional methods with machine learning models. Particularly, Gaussian process regression (GPR) models, developed specifically to forecast the daily bottom heating value of municipal solid waste (MSW), have become an important tool for increasing the efficiency of waste facilities and optimizing waste management processes. These models are critical to ensure the stable operation of waste-to-energy (WtE) plants by accounting for the change of physical and chemical properties of MSW over time.

This thesis study aims to analyze recyclable waste datasets collected from different regions and waste types by performing statistical data analyses, and then it aims to forecast their amounts by using both traditional forecasting and supervised machine learning methods. Forecasts provide information on the reuse, regeneration, and

recycling of solid wastes. In this study, "waste" is referred to as recyclable solid waste that is disposed of as by-products of packaged goods, such as cartons, glass, paper, cardboard, metals, and plastics. Thus, they are considered under two categories as paper and plastic wastes throughout this thesis. However, collecting such real waste datasets at different regions is difficult and this prevents making accurate predictions regarding their future amounts in real life. Therefore, this study is primarily aimed at making forecasts with real datasets of packaging wastes obtained from different collection points in Izmir, Turkey. Various approaches are used in the literature for waste forecasting and revenue maximization such as neural networks, Gaussian process regression, and two-level decision-making (Hoque and Rahman, 2020; Muneeb et al., 2018; Birgen et al., 2021). However, the number of studies on waste datasets is limited. Also, having forecasts for both multi-waste types and multi-regions within the same city on a same forecasting horizon makes forecasting studies more complicated. Further, although there are also some stochastic models for waste forecasts in the literature such as the study conducted by Afrinaldi (2020), obtaining a valid stochastic distribution model is also difficult for different regions and products. In addition, using and creating multivariate (or joint) distributions of the same family is impractical to model the common behaviour of different observed datasets, that are collected from different and independent location points. Rather, the approach is similar to fitting each dataset with a different univariate random variable, however, the collection of these random variables would not have a control like in a multivariate joint distribution (Law and Kelton, 2007). This is the reason each dataset is employed separately for the analysis in this study. Also, the interest in this study is in how realizations of such forecast data be generated. This is why this study aims to generate forecasts with an automated model that would be suitable for the forecasting of various waste types obtained from different regions. The forecast results are compared and verified by creating different models. Accordingly, the next part defines the problem of this study in more detail.

1.2. Problem Definition

This study aims to analyze the recycled waste amounts collected at different points and waste type in a city and then forecast them further with both traditional and machine learning approaches. The datasets are analyzed according to the daily collected data from different selected points in the city and forecasting results are obtained. There are various regions and multiple waste types in the dataset. Waste types are divided into two types: paper and plastic. In addition, machine learning techniques that are popular nowadays are also used, as well as traditional forecasting techniques in this study. The results obtained from the traditional forecasting methods are also verified with machine learning approaches. In this study, XGBoost (Extreme Gradient Boosting), ARIMA (Autoregressive Integrated Moving Average) and KNN (K-Nearest Neighbors) are used as supervised machine learning approaches. Waste forecasting results can provide valuable insights for various institutions, including opportunities to generate revenue through recycling initiatives and facilitate budget planning specifically tailored to support these efforts.

1.3. Research Methodology

General forecasting methods consist of various techniques used in the fields of statistics and machine learning. These methods include statistical and mathematical approaches used to forecast the value of a variable (Hyndman and Athanasopoulos, 1998). Accordingly, regression analysis, time series forecasting, machine learning, and Bayesian methods are common general forecasting methods in the literature. They are briefly explained as follows.

Regression Analysis is used to model the relationship of a dependent variable with independent variables. In particular, regression techniques such as linear regression and logistic regression are frequently used. Time series forecasting is beneficial to predict future values of time-dependent data. These methods include techniques such as moving average, exponential smoothing and holt-winters. Machine learning involves a set of techniques and algorithms used to identify data-driven patterns and predict future values. These methods include techniques such as decision

trees, support vector machines, random forests, gradient boosting machines and deep learning. Bayesian methods are statistical methods that make predictions by combining uncertainties and previously known information. These methods work on the basis of Bayes' theorem and include techniques such as Bayesian regression, Naive Bayes classifier, and Markov Chain Monte Carlo (MCMC) (Hyndman and Athanasopoulos, 1998).

Among all these general categories of forecasting methods, this thesis study focuses on time-series forecasting, because the collected waste datasets include the quantities of different types of waste collected on various days. These datasets consist of data obtained during waste collection operations on a daily basis. The total amounts of different types of waste, paper and plastic, are recorded for each day. Additionally, these datasets can be used to create a time series forecasting to predict future waste quantities. Time series models are statistical and mathematical techniques used to analyze and forecast time-varying data. There are different time series methods such as Holt-Winters, ARIMA (Auto-Regressive Integrated Moving Average), SARIMA (Seasonal Auto-Regressive Integrated Moving Average), GPR (Gaussian Process Regression). Among these models, Holt-Winters and ARIMA methods are particularly selected for the analysis as the main time-series methods in this study. Time series models are used to predict future waste amounts by taking into account trends, seasonal variations and other important patterns of past waste amounts. This study emphasizes how collected waste datasets can be used for time series forecasting and which time series models can be preferred.

Also, this study focuses on machine learning approaches. Since they have the ability to identify complex relationships by learning from datasets. Also, machine learning models can provide high forecast accuracy by training on datasets. It also can learn from data and continuously improve the model without human intervention (Mitchell, 1997). These models can easily adapt to different data structures and changing conditions. Finally, it is capable of processing large datasets and solving complex forecasting problems. On the other hand, machine learning includes a wide variety of algorithms, providing suitable solutions to different forecasting problems. For instance, deep learning can work on complex data structures, while K-nearest

neighbour (K-NN) can be effective on more limited datasets. XGBoost is preferred due to its ability to identify complex relationships within data, fast training times and generally very accurate forecast ability.

The differences between machine learning models compared to traditional forecasting methods can be listed as follows. First of all, machine learning builds models based on large amounts of data and extracts patterns from this data. Traditional forecasting methods are generally based on theoretical models and use predefined mathematical or statistical formulas (Hyndman and Athanasopoulos, 1998). Machine learning can deal with a wide variety of data structures and complex relationships. During the learning process, the model adapts to the data to identify relationships between the data. Traditional forecasting methods generally rely on simpler and more limited model structures and may have difficulty handling complex data structures (Hyndman and Athanasopoulos, 1998). Machine learning emphasizes that features must be carefully selected and processed to make the model perform better. In traditional forecasting methods, feature selection and processing are less important and a more limited set of features is usually used. Machine learning generally performs well on large datasets and can identify more complex relationships. However, these models carry the risk of overfitting and can affect the model's ability to generalize. Traditional forecasting methods generally work with simpler models and require less data, but may have difficulty identifying complex relationships. Machine learning models can be updated as new data is received and model performance is evaluated. Traditional forecasting methods generally remain stable models and do not require updates.

This thesis study follows a case study approach. A case study is a research method in which a researcher examines in detail a specific "case" such as an individual, group, organization, event, or situation. This method can be used to understand a complex situation, examine the causes and consequences of a particular event, or confirm a theory. The case study as a research strategy represents a comprehensive method, encompassing specific approaches to both data collection and data analysis within its design logic. Case studies are often used as a starting point for a study and can later lead to more comprehensive research (Yin, 2013). This study presents a case study based on the analysis of datasets containing the amounts of paper and plastic waste

collected from different points in the Izmir Region. The purpose of the case study is to analyse the amounts of different waste collected from different regions. Details of the case study, how the data are obtained, and the importance of forecasts in the data collection process are detailed in Chapter 3. The obtained dataset is analyzed and pre-processed for forecasting methods. These analyses are mentioned in Chapter 4. The study begins with the use of traditional forecasting methods and continues with selected supervised machine learning algorithms.

1.4. Summary of Chapters in the Study

This study is organized as follows. In Chapter 2, a survey of relevant literature is given with discussions on forecasting with machine learning approaches, mathematical models and stochastic processes. Then, Chapter 3 presents the case study of the waste collection process in detail. In Chapter 4, the methodology of the study is presented. In Chapter 5, traditional forecasting methods are studied. Chapter 6 includes testing and validation of processes with selected supervised machine learning algorithms. Then, Chapter 7, introduces a framework and a discussion for automating the forecasting framework. Finally, Chapter 8 presents the conclusions at the end.

CHAPTER 2: LITERATURE RESEARCH

In this chapter, existing studies based on forecasting via several methods on the different types of waste amounts are surveyed and summarised. Some articles on maximizing income from waste recycling are presented. Accordingly, studies in the literature can be categorized in three ways according to solution approaches. These are studies that use machine learning approaches, studies that apply mathematical models and studies that are based on stochastic processes. Each category of studies is surveyed respectively in the following.

With regard to forecasting studies that employ machine learning approaches, Birgen et al. (2021) develop the Gaussian Process Regression (GPR) model to forecast the daily low heating value of municipal solid waste using waste historical data. Unlike the static models in the literature, the GPR model can learn from new data. According to their study, good forecasts of heating value may help utilize the full capacity of the facility and reduce fluctuations in energy production. Barrow and Kourentzes (2018) evaluate the use of various univariate time series prediction methods to forecast intraday call arrivals in the existence of outliers in a call centre. In addition to statistical methods, artificial neural networks (ANNs) are emphasized. It is concluded that the neural networks method accurately predicts the call centre data and can model complex outliers. In another study, Garre et al. (2020) aim to develop tools based on data analytics to estimate the magnitude of inconsistencies in food production due to uncertainty and to increase corporate profitability. Much uncertainty exists in food production and deviations between planned production and actual outputs are emphasized. These inconsistencies cause environmental impact as well as an economic cost for the company, such as waste disposal. Machine Learning algorithms are used to predict deviations in food production and uncertainties regarding the amount of waste produced are reduced (Garre et al., 2020). Machine Learning models state that with progressive parameter selection, they have greater predictive capacity than a linear model (Garre et al., 2020). As a result, accurate forecast models can potentially increase the profitability of companies. Abdella Ahmed et al. (2022)

emphasize that the type and amount of municipal solid waste plays an important role when deciding whether solid waste should be managed and converted into investment. The article examines the impact of various lifestyles on the municipal solid waste generated. For this purpose, three different areas with different standards - poor, social and rich - are used in the Al-Basaten district in East Cairo. Waste from each region is obtained separately, classified as plastic, glass, paper, cardboard and organic waste. Additionally, forecasts are made for the amount and type of municipal solid waste generation. Al-Basaten municipal solid waste is predicted using long short-term memory (LSTM) and deep learning time series forecasting (DLTSF) network. The average amount of solid waste for poor, social and rich regions is 0.42, 0.65 and 0.86 kg/person/day, respectively. The results showed that the proposed prediction model could accurately predict future plastic, glass, paper, cardboard and organic waste series values for rich, social and poor regions. This study helps decision-makers work to derive the greatest benefit from waste recycling and allows them to determine an appropriate program to organize and benefit from waste according to the living and financial norms of each region. Oguz-Ekim (2021), analyse municipal solid waste (MSW) generation estimation which is regarded as one of the biggest challenges of integrated solid waste management systems especially for developing countries where data collection is limited. Different machine learning methods are applied to the annual total waste amounts obtained from different countries. These methods are backpropagation neural network (BPNN), support vector regression (SVR), and general regression neural network (GRNN). A comparative evaluation of these different methods is made based on gross domestic product, local goods consumption and resource productivity. Additionally, these algorithms are tested on the data sample obtained from the Ministry of Environment of Turkey. The results of this study are expected to provide a general framework for stakeholders in Turkey to develop MSW management strategies across the country, and these results can be similarly used in other developing countries. In this article, Relative Percentage Error (RelErr) and coefficient of determination (R^2) are used for the predictive performance measure of the model. As a result, it can be concluded that for Turkey and other countries, BPNN and SVR methods can be successfully applied to predict MSW

production, but BPNN is slightly better. When input and output variables are well defined, machine learning approaches can provide a good estimate of waste generation, and this projection can be used for different countries (Oguz-Ekim, 2021). Hoque and Rahman (2020) use an Artificial Neural Network (ANN) model to estimate solid wastes collected between 2012-2016. The 120-month solid waste amount is used as input data for the model. In the ANN-based solid waste prediction model, the existing weighbridge data are used with the determination coefficients of 0.85 and 0.86 for training and tests. As a result, it is stated that the proposed landfill estimation method and final disposal methods with ANN can be used as alternatives that help better planning and management of the landfill. Using a different approach, Sunayana et al. (2021) introduce an estimation method with generated using nonlinear autoregressive (NAR) models. It is aimed at estimating municipal solid waste production in Nagpur (India) for the year 2023. In conclusion, AI-based models are helpful for situations where it is difficult to run inventory due to budget constraints faced by the municipality. Twumasi and Twumasi (2022) emphasize the importance of demand forecasting due to the balance between shortage and waste in blood products, as well as uncertainty in demand. These estimates can be used to improve blood availability. The analyzed data contain missing values and outliers. For this reason, a variety of models different from the time series models are being tested. As a result, K-Nearest Neighborhood regression (K-NN) performs well in predicting blood demand.

Further approaches could include also mathematical models. With regard to their use on solid waste amounts, Muneeb et al. (2018) introduce a two-level decision-making model for municipal solid waste (MSW) management. It is aimed to minimize the net transportation cost and maximize the income from treatment facilities. Fuzzy target programming is applied to obtain a solution. The model is presented to effectively manage interactions between waste transportation, inventory holding and production planning in an uncertain environment. Xu et al. (2017) discuss the challenges posed by the design of the global reverse supply chain in recycling solid waste. They use a mixed integer-linear programming method. As results show, they may obtain a better global reverse supply chain. Razzaq et al. (2021) contribute to estimating the municipal solid waste (MSW) recycling impact on economic growth

and CO₂ emissions. It uses bootstrapping autoregressive distributed lag modelling to investigate the relationship between energy efficiency utilizing quarterly data from 1990 to 2017. As a result, one percent increase in waste recycling contributes 0.317 percent to economic growth in the long term and reduces carbon emissions by 0.209 percent.

Also, the following gives some examples on the application of stochastic models on forecasting in general. To illustrate, Tsiliyannis (2018) develops a method for real-time predicting product returns in re-manufacturing. The model allows non-stationarities, uncertainty and varying utilization of reusable returns. Monte Carlo Markov Chain (MCMC) simulation provides an assessment of the impact of the forecasting method. As a result, this study may be beneficial for increased acquisition of returns with decreased stock inventories and efficient planning of re-manufacturing operations. Afrinaldi (2020) develops a Markov chain to model a product's life cycle. The model is used to forecast significant information about the behaviour of the product throughout the life cycle. It analyzes the life cycle of plastic produced in the study. The proposed model predicts that the total energy consumption. Also, the study shows that on average plastic will visit the production, recycling, and use stages many times. Pillai and Chandrasekharan (2008) model the flow of material through the production system as an absorbing Markov chain featuring uncertainty owing to scrapping and reworking. This study defines production system parameters under scrapping and reworking and accurately predicts the quantity of raw materials required. Wu et al. (2017) present a life cycle optimization model that uses the Semi-Markov process to make decisions for bridge maintenance. The bridges in the highway system deteriorate due to reasons such as material degradation, harsh environment and increased traffic. As a result, cost savings and more efficient sustainability of bridge systems are achieved. Van der Walt and Bean (2022) study minimizing wastes generated from the caterings given on the plane. Due to the uncertainty of passengers' food preferences, in-flight caterers tend to over-produce meals. A stochastic and multi-objective mixed integer programming model is designed to define and solve this problem. The model estimates the probability distribution of the flight's total food demand based on the output of a time-inhomogeneous Markov chain forecasting model. As a result, the passenger

satisfaction level is maximised while minimizing the amount of waste generated according to the model.

Table 1. Summary Table of the Proposed Methods from Literature

	K-NN	GPR	ANN	NAR	ARDL	LSTM	DLTSF	BPNN	SVR	GRNN
Twumasi C. And Twumasi J. (2022)	x									
Birgen et al. (2021)		x								
Hoque and Rahman(2020)			x							
Kumar et al. (2021)				x						
Barrow D. and Kourentzes N.(2018)			x							
Razzaq et al.(2021)					x					
Abdella Ahmed et al. (2022)						x	x			
Oguz-Ekim (2020)								x	x	x

K-NN: K-Nearest Neighbor, GPR: Gaussian Process Regression, ANN: Artificial Neural Network, NAR: Nonlinear Autoregressive Model, ARDL: Autoregressive Distributed Lag Model, LSTM: Long Short-Term Memory, DLTSF: Deep Learning Time Series Forecasting, BPNN: Backpropagation Neural Network, SVR: Support Vector Regression, GRNN: Generalized Regression Neural Network

The studies on waste forecasting in the literature show the various uses of machine learning methods in estimating waste quantities. The algorithms preferred for these methods are summarized in Table 1. While demonstrating various applications of machine learning methods in the field of forecasting, different research approaches in the literature are also highlighted. On the other hand, bootstrapping and optimization models, and heuristic methods are also used in the literature.

The findings in these articles support the use of machine learning algorithms in forecasting waste quantities. For example, Abdella Ahmed et al. (2022) shows that Long Short-Term Memory (LSTM) and Deep Learning Time Series Forecasting (DLTSF) methods are successful in predicting the amount of solid waste in different regions. Birgen et al. (2021) show that it is effective to predict the daily lower temperature value of MSW using the Gaussian processes regression model. Garre et al. (2020) showed that ML models have better prediction capacity than linear models. Oguz-Ekim (2021) explains that BPNN and SVR methods are superior in forecasting MSW. Twumasi and Twumasi (2022), on the other hand, uses the K-Nearest Neighborhood regression model for blood demand forecast. Machine learning includes a wide variety of algorithms that provide suitable solutions to different forecasting problems. These findings support the use of machine learning techniques

in waste forecasting.

In this thesis study, the first paper and plastic recyclable waste datasets are analysed, and statistical inferences are evaluated using descriptive and correlation analyses for different regions. After that, traditional forecasting methods are applied. Then, supervised machine learning approaches are performed for improved forecasting.



CHAPTER 3: CASE STUDY

This thesis forecasts the collected recyclable waste amounts at certain times as time series. This study aims to make forecasts based on expanded data and two different types of recyclable waste types.

In this chapter, the selected case is based on the 12 locations in a particular area of the city called Izmir in Turkey. These locations include supermarkets and stores, where relatively high amounts of paper and plastic waste are produced and delivered. A waste collection company operates privately through these locations together with a software company. The software company has a customized software structure to manage the information flow and through the waste management software, it directs the waste collection among these 12 points. Accordingly, this study includes the datasets collected from these locations from July 2021 until May 2022. The steps involved in this case study are explained herein. First, waste fullness messages (i.e., signals) coming from these locations are monitored, but electronic signals are not regularly received from these points, instead, these signals are known to be sent by officials in the stores. While the waste is collected, the vehicle does not go directly to the point where the signal is received, but it proceeds after a route is created according to the incoming occupancy messages. Then, the vehicle visits waste collection points according to the generated route. At each visit, waste amounts are measured with an electric scale and the measurement value is processed digitally. The details of quantity, waste type, location and collection time are entered into the database of the waste collection software application. Waste collection takes place in this way for each location during the day. The collected waste is delivered to the waste collection centre. If the vehicle capacity is not enough for other locations, this time the vehicle continues to collect the waste by unloading it and going to other remaining locations. However, in this study, it is assumed that such details are ignored and that waste is not collected beyond the capacity of the vehicle on each route. The dataset collected in this study is visualized and transferred with certain steps specified in the flow chart in Figure 1.

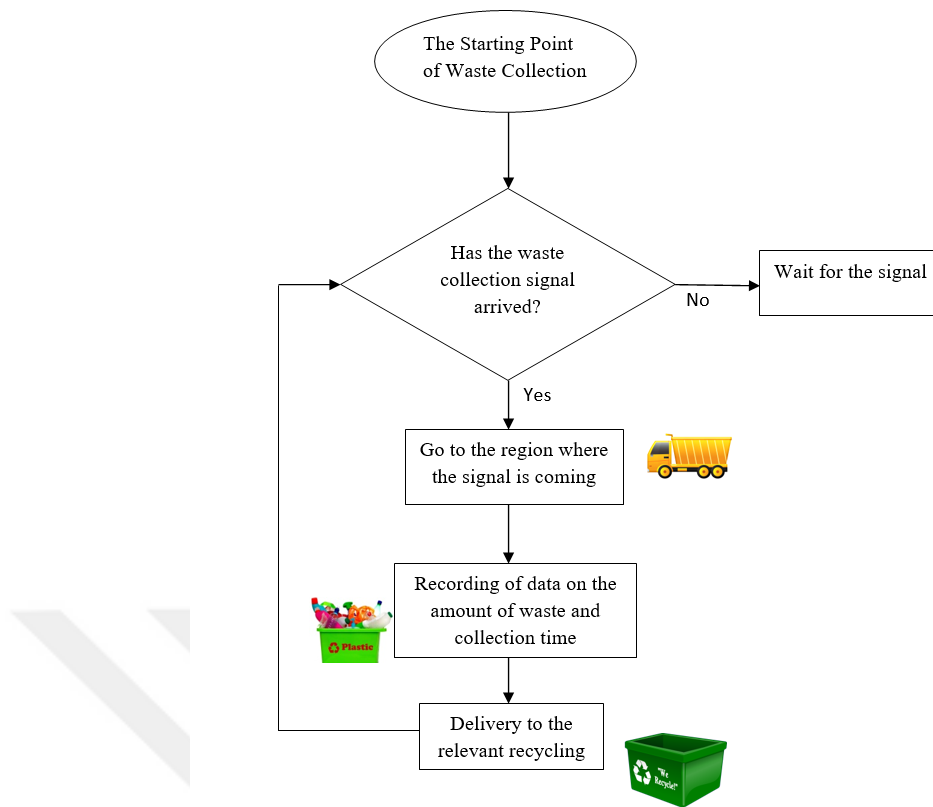


Figure 1. A Flow Chart for the Stages of Waste Data Used in This Study

The importance of the case study consists of the forecasting results of different waste datasets collected from different locations that could provide economic, environmental and time benefits to the waste collection process in real life. Specifically, a more precise waste forecast can facilitate better waste transportation and logistics planning processes. According to the forecasted waste amounts, the resources and logistics operations required for the collection and transportation of waste could be planned more effectively in terms of minimizing transportation costs and increasing the utilization of collection trucks. This would further facilitate the minimising Carbon footprint of such deliveries and also time saving. Also, budget planning can be done and inefficiency in the use of resources is reduced. More accurate waste forecasting can be used to set recycling targets and maximizing revenue from recycling.

In addition, Smart Collection systems in the future could be applied so that waste collections would be carried according to the microchip signals in the collection locations. However, considering that the use of these smart sensors in these locations may be costly, therefore the forecasting of waste amounts is important.

CHAPTER 4: METHODOLOGY

The main purpose of this study is to determine the methods that give the best waste forecast results by applying different forecasting methods for different waste types and regions. The method followed for this purpose is the supervised learning model.

The Flow chart is depicted to show the steps of the study in Figure 2. An explanation of where the data are obtained and data details are included in Section 4.1. In Section 4.2, the examination of the expanded dataset used in this study and the pre-processing phase of the data are detailed. In Section 4.3, the statistical interpretation of the dataset is explained with graphics. In Section 4.4, the correlation chart related to the dataset is interpreted.

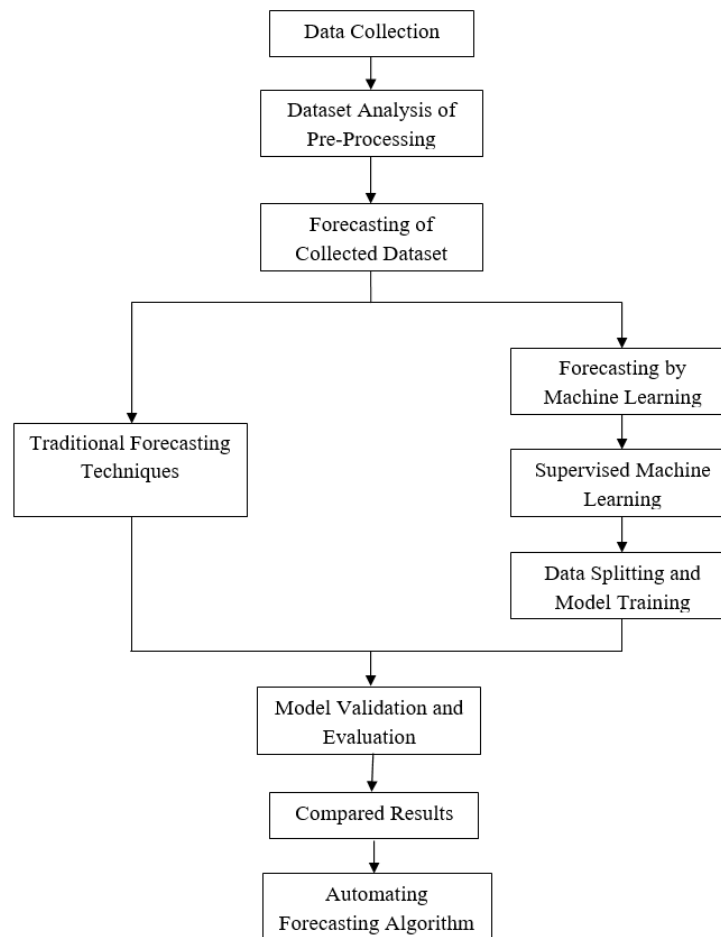


Figure 2. Flow Chart of Methodology

The methodology of this thesis begins with the collection and pre-processing of

the dataset. It continues with data analysis and then machine learning model selection. To make the forecast, selected models train and compare the performance of models. The model that gives the best forecast results according to the features of the dataset is presented.

4.1. Data Collection

First, data was taken from the official site of the Ministry of Environment and Urbanization and the Turkish Statistical Institute from 2014 to 2018 for this thesis. In this study, packaging waste observed in Turkey such as plastic, metal, composite, paper cardboard, glass and wood was used. Table 3 includes the total production, usage and recycling amounts by year. According to Table 3, 40 million metric tons (Mt) of packaging waste is produced, 20 million Mt are used, 12 million Mt are recycled, 141 million Mt are sanitary landfills, 847 thousand Mt are incinerated, 372 thousand Mt are compost and 2 million Mt are co-incinerated from 2014 to 2018.

Table 2. Packaging Waste Cumulative Collected Data Related Production, Use and Recycle in Turkey Between 2014 to 2018

Packaging Waste	Total Production	Total Use	Total Recycle
2014	7,286,184	3,948,307	2,422,521
2015	7,085,518	4,183,309	2,530,664
2016	7,773,896	3,850,712	2,226,273
2017	8,633,055	4,127,867	2,198,845
2018	8,937,233	3,893,395	2,375,517
The Overall Total	39,715,886	20,003,590	11,753,820

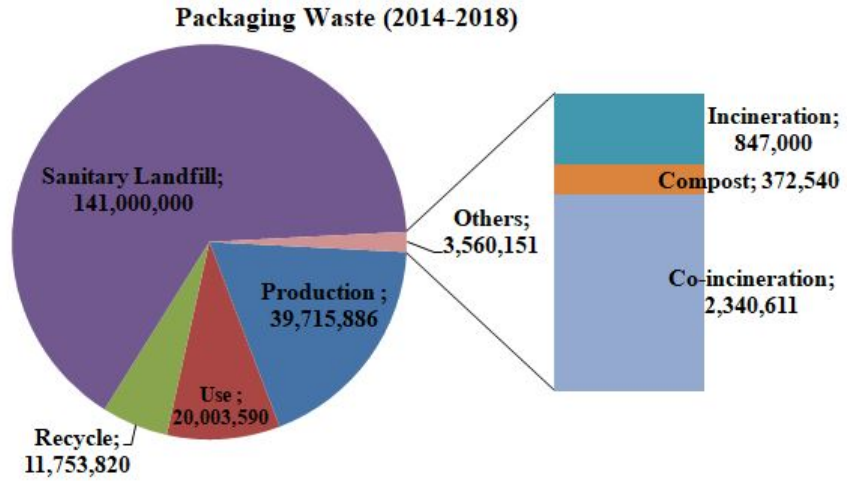


Figure 3. Total Number of Packaging Waste at All Stages Between 2014-2018

Figure 3 presents the total number of packaging waste between the years 2014 and 2018 in the following. These analyses are just sample aggregated data (yearly data) collected from the Environmental Ministry however datasets collected for the case study in this thesis from different regions are analysed systematically in subsections. Since the data received from the Ministry of Environment and Urbanization are cumulative, it was not sufficient to perform detailed forecasting analyses, this study followed a real case study of recyclable waste collection in Izmir region of Turkey by using real-time paper and plastic waste dataset obtained from a waste company.

4.2. Dataset Analysis of Pre-Processing in Real Dataset

There are 12 areas and an average of 264 paper and 238 plastic waste data are obtained from each area. The values in the dataset are collected in kg. In the data pre-processing stage, it is aimed to make the collected data suitable for the purpose of the study. Since some days in the dataset are seen as cumulative values, gaps within the month are filled. Average values are written by dividing the next day's value by the difference of the previous day. For example, on day 1 there is waste data value, and on day 3 there is cumulative value. It looks like data from day 2 was skipped. However, the data of the 3rd day is the cumulative value of the 2nd and 3rd days. Therefore, the missing data on the 2nd day is completed by averaging the 3rd day data. Additionally, there was a data gap in many regions in the dataset between

approximately 24.11.2021 and 12.12.2021. This gap was filled differently, unlike the previously mentioned method. This gap was filled with the mean substitution method. Missing data in that region is filled by averaging the other data in the area where the dataset is located. Table 3 presents the total amount of waste collected from each region regarding paper and plastic waste. In Figure 4 the total paper waste data and in Figure 5 the total plastic waste amounts are shown on the histogram according to different regions.

Table 3. Total Amounts of Paper and Plastic Waste Collected in Each Region

Region Name	Total Amount of Paper Waste (Kg)	Total Amount of Plastic Waste (Kg)
56906 [A]	16.711	860
57062 [B]	19.305	3.685
56964 [C]	17.420	884
56905 [D]	42.796	1.963
57138 [E]	26.484	1.314
56959 [F]	17.359	326
56904 [G]	10.157	434
56976 [H]	12.126	661
56903 [I]	9.776	385
57072 [J]	4.761	479
57075 [K]	14.832	1.071
57165 [L]	55.560	251

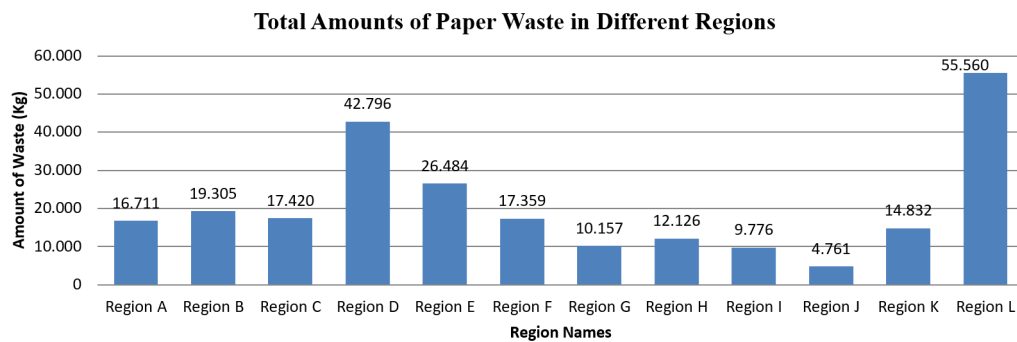


Figure 4. Total Amounts of Paper Waste by Regions

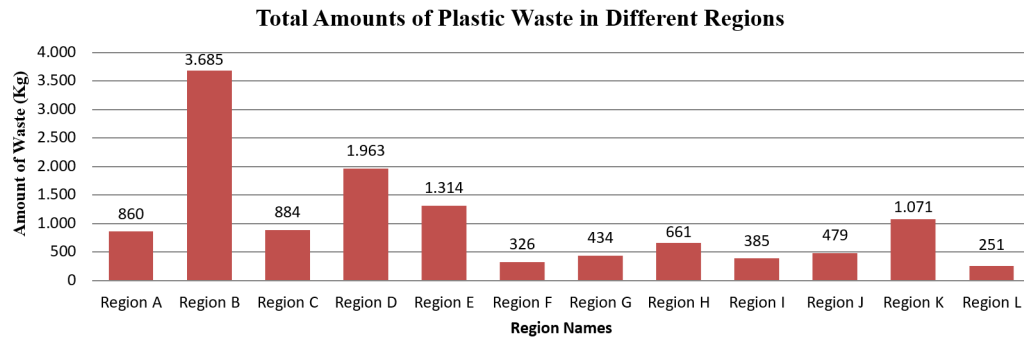


Figure 5. Total Amounts of Plastic Waste by Regions

40 weeks of collected waste amount data is shown grouped into approximately 10 months. The monthly collected amounts from Paper and Plastic waste data are shown in the figure 6.

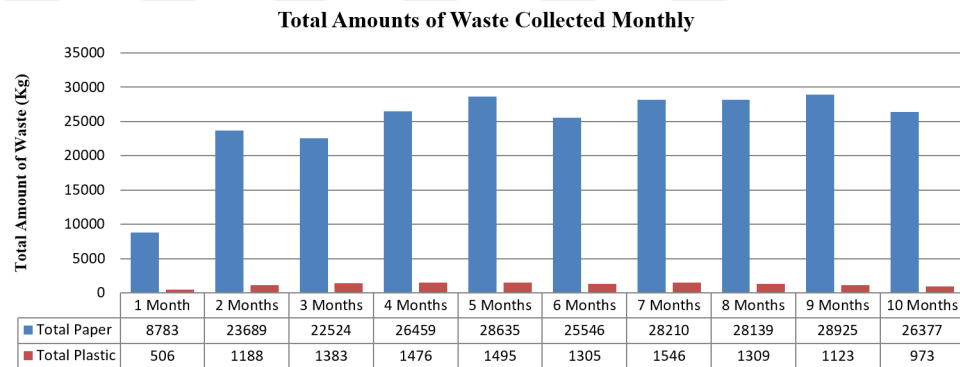


Figure 6. Total Amounts of Waste Collected Monthly

Paper and plastic waste data is collected daily. These waste data consist of discrete time series data. When examining the daily data type, the days of the week are interpreted with box plots. In Figures 7 and 8, there are the distributions of average paper and plastic waste data by days of the week, collected from 12 regions. Box plots simultaneously present various statistical properties of the dataset (e.g. minimum, maximum, median, interquartile range, etc.) and help us quickly understand the overall distribution of the dataset. For example, the following statistical properties are presented on paper waste values for Monday. The median waste amount for Monday is 67.0, with the first quarter (Q1) being 55.75 and the third quarter (Q3) being 76.0. The waste quantity range is from 40 to 122. It seems that the days with the highest amount are Tuesday and Thursday for paper waste. This is because the median values of these

days are higher than other days. Specifically, the median value of waste amount on Tuesday is 94, while on Thursday it is 92. The median values of other days are lower. According to plastic waste data, the days in Figure 8 where the highest amount of waste is collected can be interpreted as Tuesdays and Wednesdays. These are determined because the median value of both is higher than other days. The median value of waste amount on Tuesdays and Wednesdays is 4.5. On other days, the median value appears to be 3 and 4. Outliers are found on Thursday, Friday and Sunday. Looking at the maximum values for the amount of waste on each of these days, it is seen that it is 8.5 kg on Thursday, 7.9 kg on Friday and 7.6 kg on Sunday. These values exceed the maximum waste amounts observed on other days, indicating the presence of outliers on these specific days.

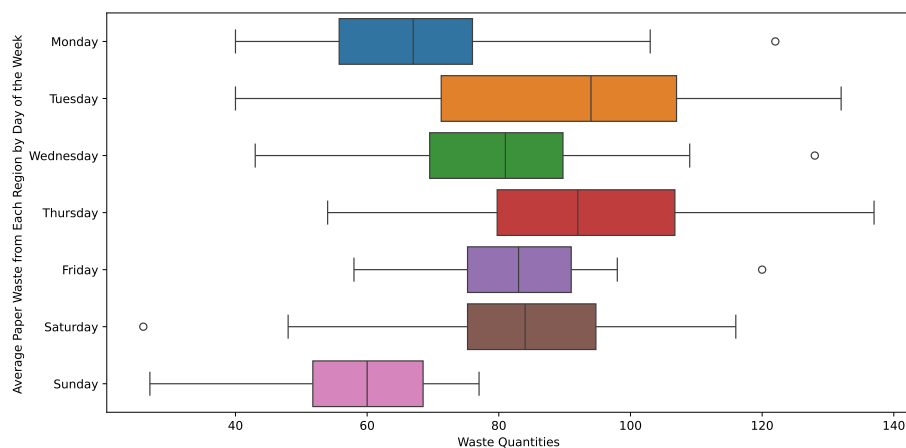


Figure 7. Distribution of Average Paper Waste Data by Days of the Week

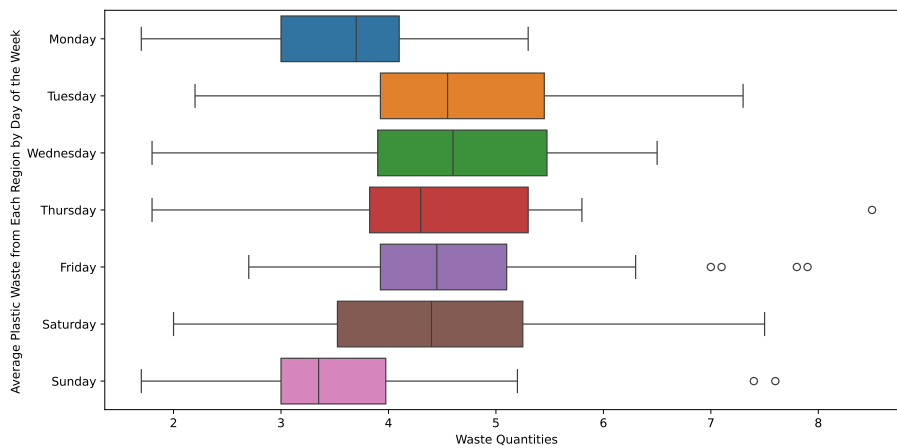


Figure 8. Distribution of Average Plastic Waste Data by Days of the Week

4.3. Analyses of Descriptive Statistics of Waste Datasets

The twelve regions have different region names as defined inventory ID (example: 56906). Statistical data for each inventory ID and waste type are also shown in Tables 4 and 5. The distribution and pattern of the data could be observed with descriptive statistical analyses such as average values, minimum and maximum values, range, mean absolute deviation, coefficient of variation, skewness and standard deviation of the paper and plastic dataset.

Table 4. Statistical Analysis of Paper Waste for Each Region

Inventory ID	56906 [A]	57062 [B]	56964 [C]	56905 [D]	57138 [E]	56959 [F]	56904 [G]	56976 [H]	56903 [I]	57072 [J]	57075 [K]	57165 [L]
Mean	59,7	81,5	63,8	155,1	105,5	64,5	36,8	44,4	35,5	18,7	58,2	219,6
Min	6	10,6	13,9	29	5	12	6	4	3	1	6,4	11
Max	134	180	188	397	285	256	125	127	136	79	242	534
Range	128	169	174	368	280	244	119	123	133	78	236	523
Count	280	237	273	276	251	269	276	273	275	254	255	253
Mean Absolute Deviation (MAD)	19	27	30	48	46	30	17	13	10	9	22	90
25%	44	59	34	110	67,5	35,5	19,8	34	26	10	39,5	144
75%	73,5	106	80	187,3	141,5	82,5	45	53	44	25	71,5	288
Standart Deviation	24	33	38	65	56	39	23	18	14	13	31	114
Variance	579	1083	1424	4181	3135	1546	537	308	196	156	957	12996
Coefficient of Variation	40%	40%	59%	42%	53%	61%	63%	40%	39%	67%	53%	52%
Lexis Ratio	0,12	0,14	0,08	0,11	0,06	0,13	0,13	0,10	0,12	0,09	0,10	0,06
Skewness	0,4	0,0	1,1	0,7	0,3	1,3	1,3	1,1	1,4	1,3	1,7	0,3

Table 4 shows the statistical analysis results regarding paper waste. For example, when the results for region A are examined, it is seen that the average amount of paper waste is 59.7 kg. The lowest and highest amounts of paper waste indicate the range of

the dataset. While the lowest amount of paper waste in Region A is 6 kg, the highest amount is recorded as 134 kg. Range refers to the difference between minimum and maximum values. This shows the breadth of the data distribution. The amount of paper waste in this region is determined as 128 kg. The count represents the sample size of the waste in each region. This refers to the number of observations in the dataset. For example, in region A, there are 280 observations of paper waste. Quartiles (25% and 75%) refer to the lower and upper quartiles of the dataset. These values show the distribution of the dataset and can be used to determine whether there are outliers. Standard deviation measures the variation of data points around the mean. Higher standard deviation values indicate that data points are more dispersed around the mean. The coefficient of variation expresses the ratio of standard deviation to the average value. A higher coefficient of variation indicates that the variance in the dataset is higher than the mean value. In region A, this value is achieved by 40%. Lexis ratio measures variability over a specific time period. This can be used to analyze changes in the dataset over time. For example, looking at the ratio of a 30-day waste collection amount to the entire waste collection amount, this ratio is obtained as 0.12 in region A. Skewness measures the symmetry of the data distribution. Positive skewness indicates a bias to the right, while negative skewness indicates a bias to the left. It can be seen that there is a positive skewness of 0.4 in region A.

Table 5. Statistical Analysis of Plastic Waste for Each Region

Inventory ID	56906 [A]	57062 [B]	56964 [C]	56905 [D]	57138 [E]	56959 [F]	56904 [G]	56976 [H]	56903 [I]	57072 [J]	57075 [K]	57165 [L]
Mean	3,1	14,6	3,2	7,0	7,1	2,1	1,6	2,7	1,4	2,0	4,3	1,6
Min	1	3,0	1,0	1	1	1	1	0,2	0,2	0,5	0,5	1,5
Max	17	35	11	23	39	9	6	11	6	8	19	15
Range	16	32	10	22	38	8	5	10,8	5,8	7,5	18,5	13,5
Count	279	253	273	279	186	154	272	247	271	240	247	157
Mean Absolute Deviation (MAD)	1	4	1	4	5	1	0	1	1	0	2	0
25%	2	11	2	3	2	1	1	1	0,5	2	2	1,5
75%	3,1	17	4	10	10	2,1	1,6	3	2	2	6	1,5
Standart Deviation	1,8	5,5	1,7	4,4	6,3	1,6	0,7	2,0	1,0	0,8	3,0	1,1
Variance	3	30	3	20	40	3	1	4	1	1	9	1
Coefficient of Variation	60%	38%	52%	63%	89%	77%	46%	74%	71%	40%	68%	67%
Lexis Ratio	0,12	0,12	0,09	0,04	0,04	0,19	0,08	0,15	0,08	0,13	0,18	0,18
Skewness	3,8	0,8	1,1	0,8	2,1	2,0	2,4	2,0	1,3	2,9	1,2	12,5

Table 5 represents the statistical analysis results regarding plastic waste. For example, when the results for region A are examined, it is seen that the average amount of plastic waste is 3.1 kg. The lowest and highest amounts of plastic waste indicate the

range of the dataset. While the lowest amount of plastic waste in Region A is 1 kg, the highest amount is recorded as 17 kg. The range amount of plastic waste in this region is determined as 16 kg and there are 279 observations of plastic waste. In region A, the coefficient of variation is measured by 60%. Looking at the ratio of a 30-day waste collection amount to the entire waste collection amount, the lexis ratio is obtained as 0.12 in region A. It can be seen that there is a positive skewness of 3.8 in region A.

A boxplot is used to analyze the distributions, central tendencies, spreads and outliers by region of the dataset specified in Figures 9 and 10 below. Among the regions where paper waste is collected, the 57165 (L) region appears to be the region with the highest spread and outlier values. The variability in the amount of waste collected in region J is less than in the other regions in paper waste. According to Figure 10, in regions where plastic waste is collected, the 57062 (B) and 57138(E) regions appear to be the regions with the highest spread and outlier values.

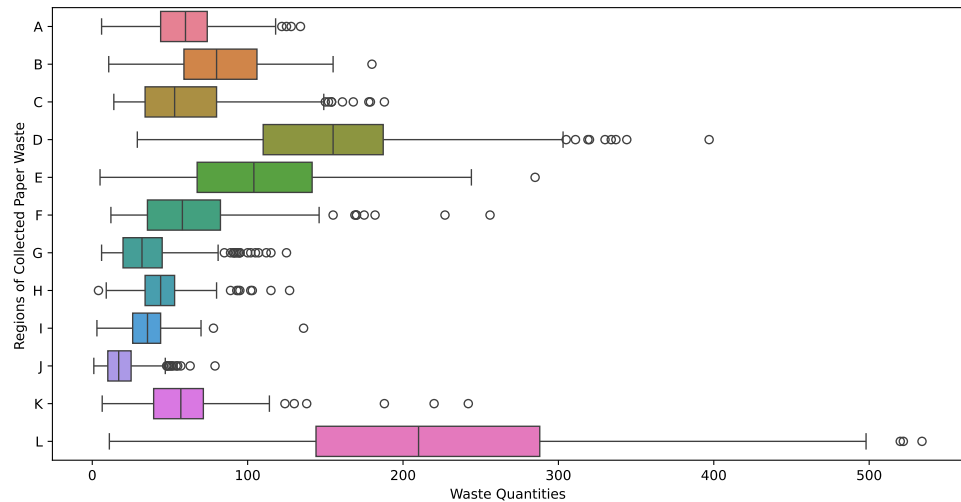


Figure 9. Box Plot of Paper Waste by Regions

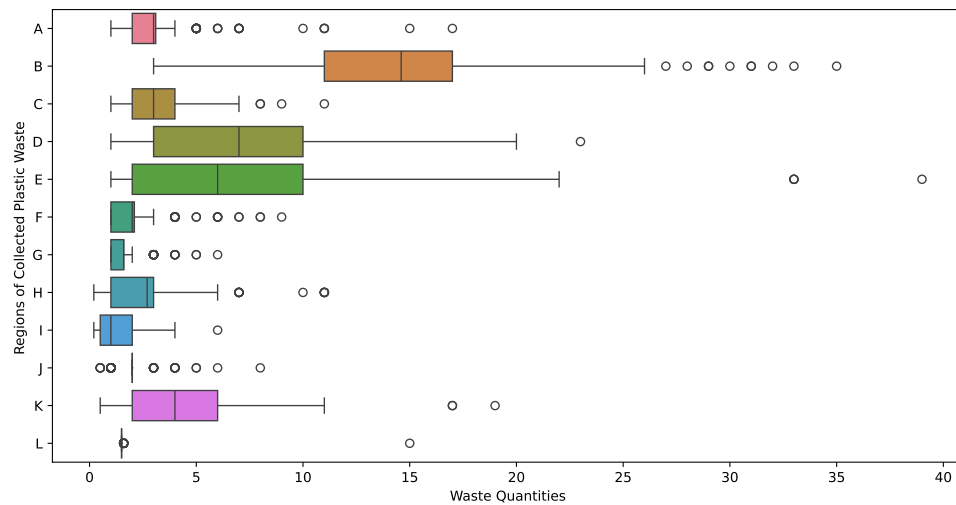


Figure 10. Box Plot of Plastic Waste by Regions

In Figures 11, there are graphs showing histograms and curve-fitting lines regarding paper waste for four regions. In regions C, F, G, J, K and L, the graphs can be interpreted as right-skewed. This implies that the peak of the graph is on the left side of the centre. Mean is greater than the median. The normal distribution is observed in other regions. The mean, median, and mode are approximately equal. In Figures 12, there are graphs showing histograms and curve-fitting lines regarding plastic waste for four regions. Paper and plastic waste graphs for other regions are in Appendix A.

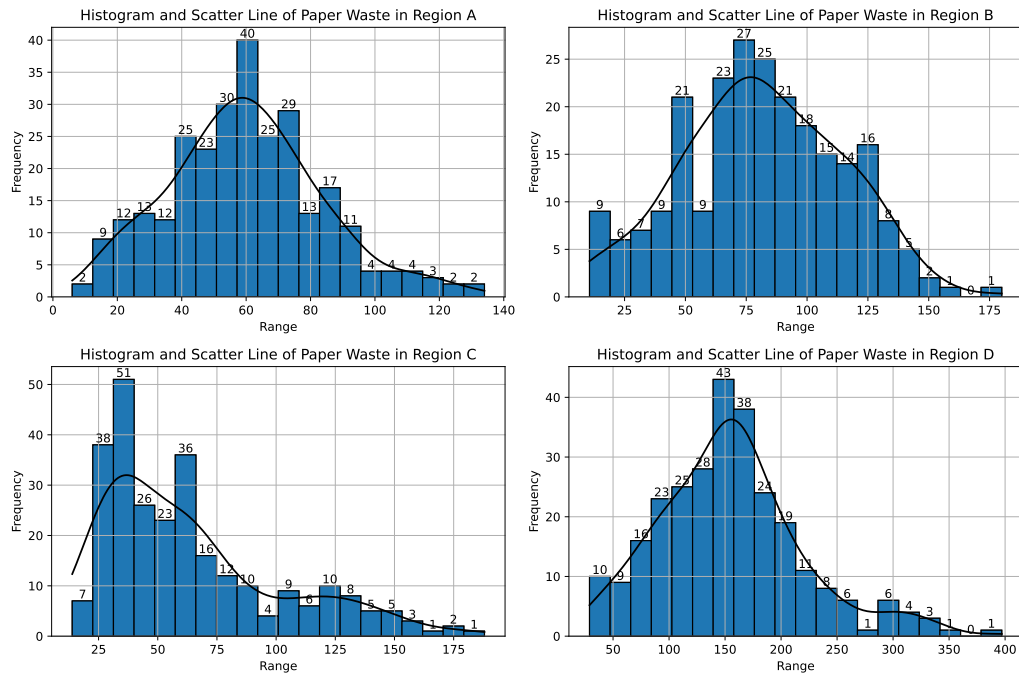


Figure 11. Histogram and Curve Fitting Line of Paper Waste in Regions

Table 6. Fit Distributions and Parameters of Paper Waste for Each Region

Region Name	Distribution	Parameter (p)
Region A	Geometric	0.01646
Region B	Geometric	0.01212
Region C	Geometric	0.01542
Region D	Geometric	0.00641
Region E	Geometric	0.00939
Region F	Geometric	0.01525
Region G	Geometric	0.02639
Region H	Geometric	0.02203
Region I	Geometric	0.02734
Region J	Geometric	0.05058
Region K	Geometric	0.01691
Region L	Geometric	0.00453

Table 6, these results show which distribution of paper waste is appropriate according to regions. For example, it shows that geometric are suitable for region A. The geometric distribution models the probability of a particular event occurring for the first time. The parameter of the geometric distribution was determined as p

is 0.01646 in region A. This refers to the probability of a certain number of events occurring within a certain period of time per unit of paper waste data in "Region A". For example, it can be interpreted as the probability of a certain number of paper waste events occurring in each unit of time. These parameters help us better understand the distribution of paper waste data and the probability of events in regions.

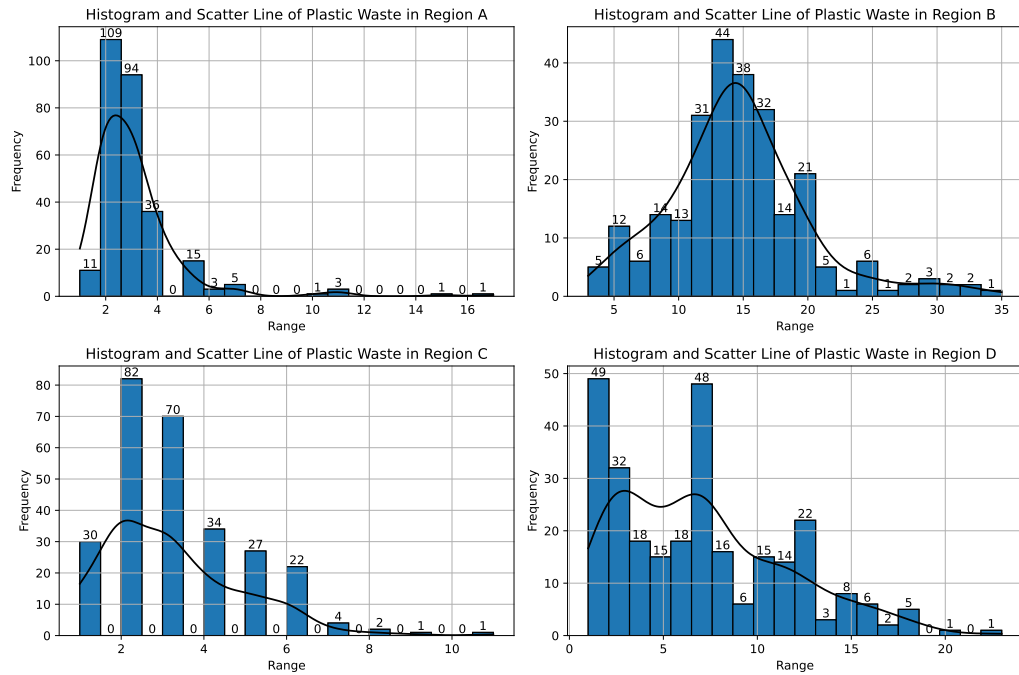


Figure 12. Histogram and Curve Fitting Line of Plastic Waste in Regions

Table 7. Fit Distributions and Parameters of Plastic Waste for Each Region

Region Name	Distribution	Parameter (p)	Distribution	Parameter (n, p)
Region A	Geometric	0.24538		
Region B	Geometric	0.0641		
Region C			Binominal	22 , 0.1434
Region D	Geometric	0.12444		
Region E	Geometric	0.124		
Region F	Geometric	0.32218		
Region G			Binominal	2 , 0.6667
Region H	Geometric	0.26502		
Region I			Binominal	3 , 0.38887
Region J			Binominal	2 , 0.68608
Region K	Geometric	0.18712		
Region L			Binominal	4 , 0.48647

Table 7, these results show which distribution of plastic waste is appropriate according to regions. For example, it shows that geometric is suitable for region A. The parameter of the geometric distribution is determined as p is 0.24538. This refers to the probability of a certain number of events occurring within a certain period of time per unit of plastic waste data in region A. For example, it can be interpreted as the probability of a certain number of plastic waste events occurring in each unit of time. Binomial distribution fit in regions C, G, I, J and L. The binomial distribution is used to model the probability of a given number of successes in a given trial or period.

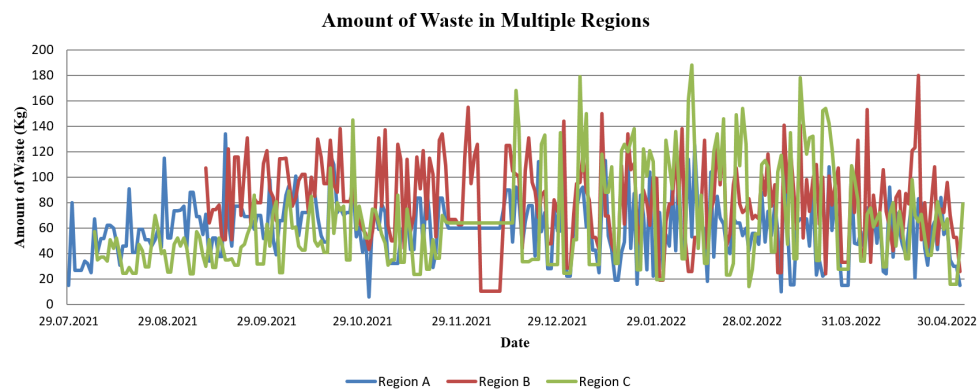


Figure 13. Amount of Paper Waste in Multiple Regions

Figure 13 shows the time-dependent paper waste collection amounts for 3 regions. There is a wide range of collection quantities, from approximately a minimum of 6 kg to a maximum of 188 kg. The peak point in region A is experienced on 16.09.2021 with 134 kg. The mean value of region A is 60 kg. The calculated peak factor of approximately 2.25 indicates that the maximum waste amount is approximately 2.25 times higher than the average waste amount. The peak point in region B is experienced on 21.04.2022 with 180 kg. The mean value of region B is 81 kg. So, the maximum waste amount is approximately 2.21 times higher than the average waste amount. Also, the peak of waste data in region C is 188 kg. The maximum waste amount is approximately 2.95 times higher than the average waste amount.

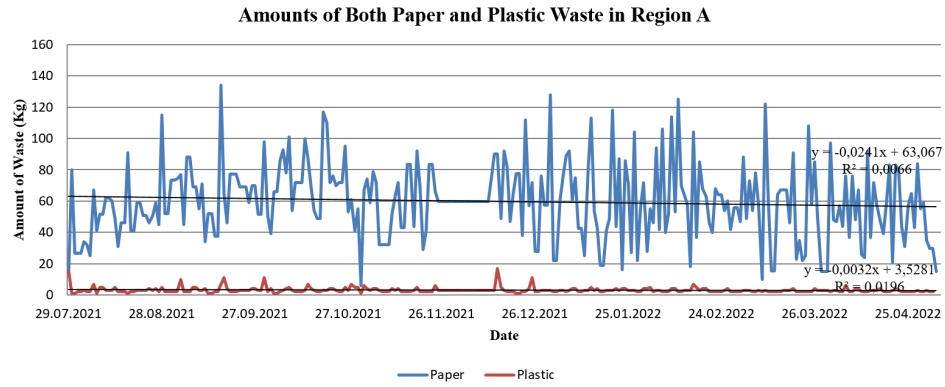


Figure 14. Amount of Both Paper and Plastic Waste in Region A

Figure 14 shows the amount of both paper and plastic waste in region A. While the amount of plastic waste is collected in the range of 0 to 20 kg in this region, it is seen that the amount of paper waste is collected up to 140 kg. According to Figure 14, a downward trend line is seen in the paper and plastic waste data, indicating that there is a decreasing trend for this dataset. Negative slope values for both types of waste indicate that the amount of waste decreases over time. R-squared values measure how well the trend line fits the data points. The R-squared value takes a value between 0 and 1, and the closer it is to 1, the better the trendline fits the data points. While the R-square value calculated for paper waste data is 0.0066, it is 0.0196 for plastic waste data. Since the R-squared values are quite low in this case, the trend line fits the data points very poorly. Low R-squared values mean that it cannot explain most of the variability in the dataset. In this case, there may be uncertainties about how reliable the trend line will be in future predictions.

4.4. The Correlation Chart

Correlation analyzes are an essential tool for understanding and modeling relationships between data. When working with time series data, the strength and nature of relationships between variables can affect the accuracy of future forecasts. In autoregressive models such as ARIMA, the performance of these models depends directly on the correlations between the data as past values are used to predict future values. Therefore, correlation analyzes are vital for the effective use of autoregressive models such as ARIMA. The chart in Figure 15 is intended to explain

the relationship between the regions. The Pearson correlation coefficient is used to evaluate the relationship between the amounts of waste collected in 12 different regions. Considering the time-dependent collection amounts and waste types of the 12 regions, the relationship between some regions is more stronger. For example, the relationship between regions A and D has a higher correlation coefficient than other regions and a positive correlation with a value of 0.54 in the graph. Also, regions E and L have a higher correlation coefficient with a value of 0.55. If the values increase towards 1 and the colour turns towards red, a positive strong relation between regions is mentioned. The values nearest to 0, mean that the relation is too weak between regions. The correlation coefficient decreases towards darker shades of blue. If the values decrease towards -1 and the colour turns blue, a negative strong relation between regions is mentioned. For example, regions D and H have a negative strong relation with -0.14.

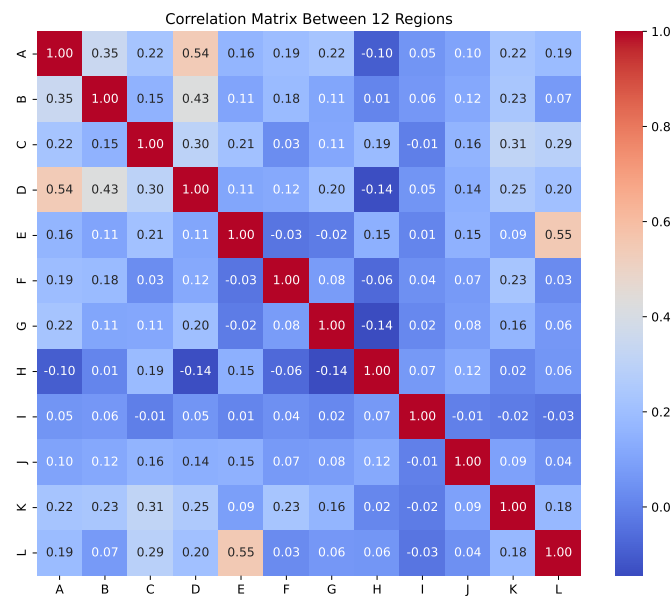


Figure 15. Correlation Matrix Between Regions

4.5. Concluding Remarks

In this chapter, data analyses are performed according to waste types for each region. In these analyses, variations and correlations are examined and interpreted.

It is observed that there is a different distribution and parameter for each region and waste type according to different waste amounts. Therefore, the results of each region present different characteristics and data features. During the forecasting analysis, the distributions of regions could be used, however, it is noted that obtaining forecasts based on distributions would not be based on the history of the wastes, and they would not present the characteristics of the history of datasets. Such a history of datasets shows time-series analyses of correlated stochastic data samples. Therefore, methods for using time-series analyses are selected and applied in this study. In this thesis, it is emphasized that the correlation and data behaviour in the data are important since the forecasting problem depends on the time series of the data.



CHAPTER 5: APPLICATION OF TRADITIONAL FORECASTING METHODS

This chapter involves testing traditional forecasting methods on the dataset before the application of machine learning approaches and representing their results. The performance metrics used for comparing forecast results are introduced in Section 5.1. Traditional forecast techniques are detailed in sections 5.2, 5.3, 5.4 and 5.5. The results of the algorithms are compared with RMSE, MSE and MAPE values in section 5.6.

5.1. Model Performance Evaluation

This section introduced the performance measures used to evaluate the forecasting methods since the performance of each method is evaluated according to these measures in this Chapter and Chapter 6. The RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error) and MSE (Mean Squared Error) deviations are commonly utilized to evaluate the inconsistency between the forecasted values by a model and the observed values represented (Das et al., 2024). The dataset is trained using several algorithms, its prediction performance is evaluated. The algorithm performance is assessed based on MSE, RMSE and MAPE metrics shown in equations 1, 2, 3 respectively. The total number of data points are shown as ‘ n ’, prediction values as ‘ $\hat{y}_i$ ’, and actual value as ‘ y_i ’ (Das et al., 2024).

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (1)$$

MSE (Mean Squared Error) is a metric that measures how far forecasted values are from actual values. It measures how much the forecasted value for each observation differs from the actual value, squares these differences, and averages them. Therefore, it is the average of the squares of the differences between the actual value and the forecasted value for each observation.

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}} \quad (2)$$

RMSE (Root Mean Squared Error) shares the same basis as Mean Squared Error (MSE) but provides one more measure by taking the square root in the final step. It provides an express the mean square error value on the same scale as the unit of actual values.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (3)$$

MAPE (Mean Absolute Percentage Error) is the absolute value of the difference between the actual and forecasted value for each observation, divided by the actual value, and the average of this ratio expressed as a percentage is taken. This measures how much the forecasted values differ from the actual values in percentage terms.

5.2. *Moving Average Method (MA)*

First, the Moving Average (MA) method is explained as one of the main traditional forecasting methods. It is applied for each region given in the case study. In this method, the forecast of the next period is found by taking the average of the collection amounts realised in the last n periods. The moving averages are calculated with the $MA(n)$ formula in equation 4 given below (Mazumder et al., 2023). n means the number of total periods. X_n is the amount of waste at period n . First, the 3-day moving average ($MA(3)$) is taken to find the forecast values and the error values between the actual value and the forecast value are examined. When choosing the " n " value, values between 1-8 are tried sequentially starting from 1 with a step increase of 1. That is, 8 different n values are chosen to determine the best and sufficient n value for this method. The most appropriate n value is found for a 3-day period, and then a 6-day period, which is twice this period, is added to the results for comparison to show that the selected n value is the best. Because the collected waste dataset is a daily time series. Therefore, the 6-day moving average corresponds to approximately weekly average waste values. However, since changes in weekly waste amounts cannot be observed in weekly periodic moving averages, it is most appropriate to choose a period value less than 6 and greater than 1, and this value is found as 3 days. The value 6 is used as a benchmark for comparison. When the graph in figure 16 below is examined,

the waste amount of a region is compared with 3 and 6-day moving averages. The fluctuation in the 6-day moving average appears to be less. The larger the window width, the more observations are used when calculating the average value, and this allows a smoother and slower changing average to be obtained. Enlarging the value of "n" can move away from the real value, while keeping it small may cause overfitting. Overfitting means when a statistical or machine learning model perfectly matches its training data (much better than the testset) but fails to generalise and wrong forecasting results (Wang et al., 2023). For this reason, values 3 and 6, which are relatively better and worse results than the others, are chosen compared to the MSE values. Paper and plastic waste types of 12 regions are examined.

$$MA = \frac{X_1 + X_2 + X_3 + \dots X_n}{n} \quad (4)$$

For moving average calculation, the "np.convolve()" function of the NumPy library, which is widely used in the Python programming language, is used. This function is used to calculate a moving average over a data series over a specific window size. To evaluate the results obtained, the results are converted into a data frame by using the "pd.DataFrame()" function of the Pandas library.

Figures 16 and 17 contain graphs showing the MA(3) and MA(6) results of paper and plastic waste in Region A. The other regions' graphics can be seen in Appendix B.

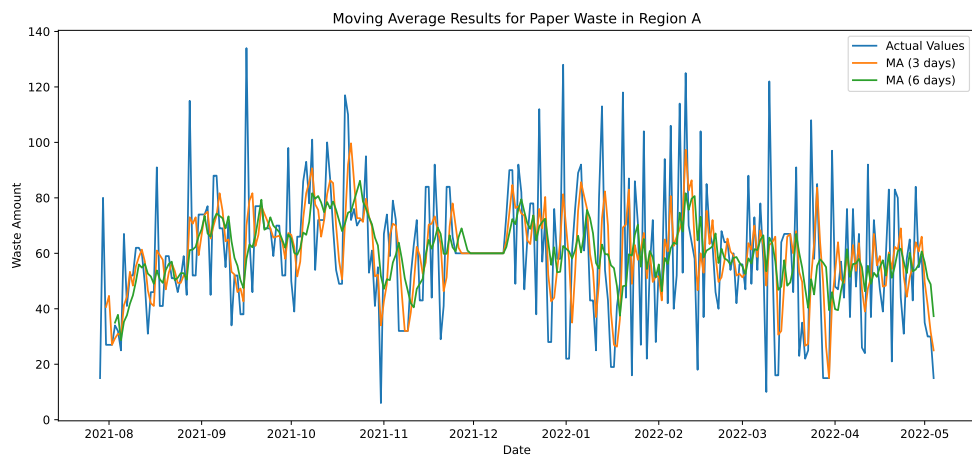


Figure 16. Moving Average Method Using 3 & 6 Days Period in Region A

In Figures 16, a moving average forecast of paper waste is calculated for region A. 3-day and 6-day periods are applied between August 2021 and May 2022. While the blue line represents actual data, the yellow and green lines represent 3- and 6-day periods, respectively. It is seen that this method cannot capture the peak points in the dataset.

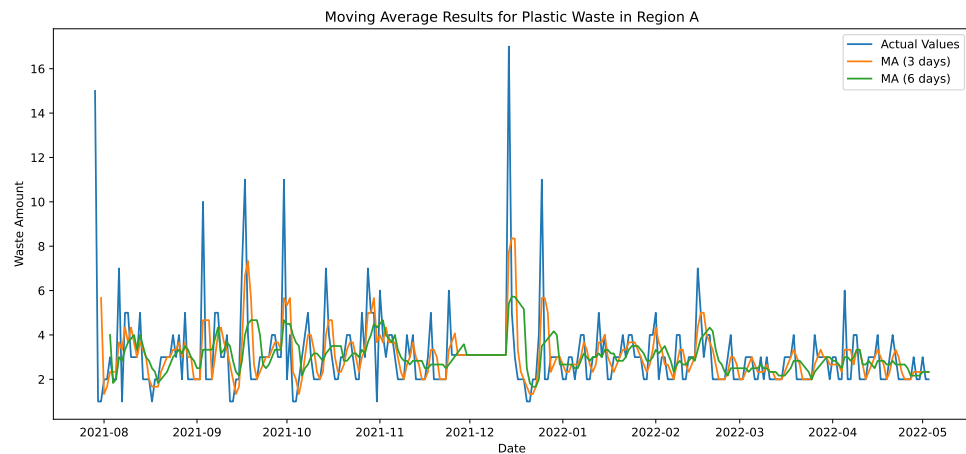


Figure 17. Moving Average Method Using 3 & 6 Days Period in Region B

In figures 17, a moving average forecast of plastic waste is calculated for region A. 3-day and 6-day periods are applied between August 2021 and May 2022. While the blue line represents actual data, the yellow and green lines represent 3- and 6-day periods, respectively. The forecasted values for the plastic waste type are closer to the actual values because the plastic waste amounts are lower than paper.

Table 8. Moving Average Method Forecast Results for Paper and Plastic Waste

MOVING AVERAGE		PAPER WASTE						PLASTIC WASTE					
Method Parameters		3 Days			6 Days			3 Days			6 Days		
		MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE
REGION A		357	32%	19	460	40%	21	1.9	30%	1.4	2.6	36%	1.6
REGION B		615	33%	25	920	47%	30	16	22%	4	21.1	27%	4.6
REGION C		765	39%	28	1.152	56%	34	1.6	38%	1.3	2.3	50%	1.5
REGION D		2.532	32%	50	3.321	39%	58	7.7	33%	2.8	10.8	47%	3.3
REGION E		1.527	41%	39	2.032	59%	45	16.1	43%	4	20.6	65%	4.5
REGION F		810	38%	28	1.079	50%	33	0.99	23%	1	1.7	44%	1.3
REGION G		317	51%	18	417	62%	20	0.25	13%	0.5	0.38	20%	0.6
REGION H		172	27%	13	234	34%	15	0.78	22%	0.9	1.48	36%	1.2
REGION I		108	27%	10	135	33%	12	0.4	41%	0.6	0.69	68%	0.8
REGION J		97	76%	10	127	107%	11	0.39	15%	0.6	0.55	24%	0.7
REGION K		472	35%	22	715	50%	27	2.02	32%	1.4	3.47	50%	1.9
REGION L		5.885	38%	77	8.180	57%	90	0.13	2%	0.4	0.03	1%	0.2

According to the results in Table 8, regions I, H and A for paper waste give better results with lower MAPE values than other regions. For regions I and H MAPE values are 27% and region A is 32%. In comparison between regions according to MAPE value, region J gives the worst results in paper waste with 76%.

Regions L, G and J for plastic waste give better results with lower MAPE values than other regions. Obtained from regions MAPE values are 2%, 13% and 15%, respectively. In comparison between regions according to MAPE value, region E gives the worst results in plastic waste with 43%.

5.3. Weighted Moving Average Method (WMA)

The weighted moving average (WMA) method is one of the applications in time series analyses. This method is studied to smooth out fluctuations in data series or to better capture a particular pattern in the data. The weighted moving averages are calculated with the $WMA(n)$ formula in equation 5 given below (Perry, 2010). n means the number of total periods. X_n is the amount of waste at period n . W_n is the weight at period n . The sum of the given weights must be equal to 1 as shown in equation 6 and larger weights are given to more recent data (Perry, 2010).

$$WMA = X_1 * W_1 + X_2 * W_2 + X_3 * W_3 + ... X_n * W_n \quad (5)$$

$$\sum_{i=1}^n W_i = 1 \quad (6)$$

Since waste data is collected daily, a 7-day moving average is calculated. The other reason for calculating a 7-day moving average is that the domain expert who shared the dataset stated that a weekly review would be more effective according to waste collection routines. In this analyses, Pandas library is used for data manipulation and data analysis. NumPy library (np.convolve() function) is used to calculate the weighted moving average and matplotlib.pyplot library is used to visualize the results. First of all, equal weights were given to the days of the week. Experts who shared the dataset deemed it appropriate to give more weight to certain days of the week. It is observed that the MSE value is lower when different weights are given to the days of the week rather than equal weights. The weights used in the graphs below are determined as (0.03, 0.06, 0.09, 0.16, 0.19, 0.22, 0.25). The following criteria are taken into account when determining these values.

- Domain expert knowledge is used to determine these values. Accordingly, for each week, the importance of a day is increasing as it gets to the end of the week. This was the first criterion considered for determination of weights.
- The second criterion was the relative importance of each day in comparison to the previous day. This is also discussed and set by the domain expert knowledge. Accordingly, days 5, 6 and 7 are the most important days and they have slightly increasing importance. Also, they have more than double importance in comparison to day 3. This is why the importance of day 5 starts with 0.19 and the other weights increase slightly for day 6 and day 7. It is noted that the exact values are selected with preliminary analysis and trials of these weights.
- Similar to the second criteria, days 1, 2 and 3 are considered as the least important days. Particularly, day 1 is the least important day among all and this is also in accordance with the increasing weights of the days given in the first criteria, and day 2 has double importance according to the relative comparison

of day 1. Also, day 3 has triple importance according to day 1.

- Overall, the importance of days can be evaluated based on day 1. Accordingly, day 2 has double importance, day 3 has triple importance than day 1, and day 4 has slightly more than five times more importance than day 1. Similarly, days 5, 6 and 7 are slightly six, seven and eight times more important than day 1. It is noted that only the day 4 importance represents a jump from four times to five times more importance than day 1. The summation of all weights must be equal to 1. This is determined by the preliminary trials and exact values of these analyses.

Figures 18 to 19 contain graphs showing the WMA(7) results of paper and plastic waste in Region A. The other regions' graphics can be seen in Appendix C.

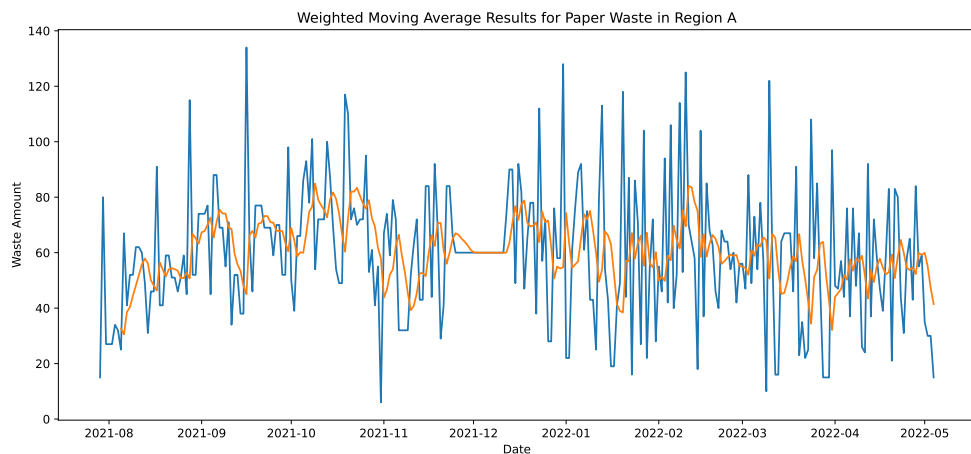


Figure 18. Weighted Moving Average for Paper Waste in Region A

In Figures 18, a weighted moving average forecast of paper waste is calculated for region A. 7-day periods are applied between August 2021 and May 2022. While the blue line represents actual data, the yellow line represents 7-day periods. It is seen that this method can more capture the peak points in the dataset than the moving average.

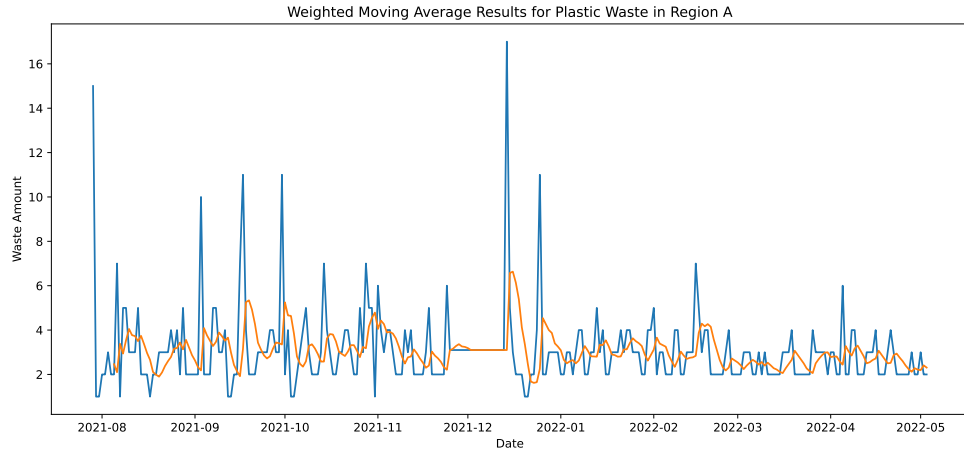


Figure 19. Weighted Moving Average for Plastic Waste in Region A

In figures 19, a weighted moving average forecast of plastic waste is calculated for region A. 7-days periods are applied between August 2021 and May 2022. While the blue line represents actual data, the yellow line represents 7-day periods. The forecasted values for the plastic waste type are closer to the actual values because the plastic waste amounts are lower than paper. However, the peak points are still not achieved.

Table 9. Weighted Moving Average Method Results for Paper and Plastic Waste

WEIGHTED MOVING AVERAGE	PAPER WASTE			PLASTIC WASTE		
Method Parameters	7 Days			7 Days		
	MSE	MAPE	RMSE	MSE	MAPE	RMSE
REGION A	385	47%	19.62	2.39	45%	1.55
REGION B	715	60%	26.74	17.08	33%	4.13
REGION C	856	62%	29.26	1.73	56%	1.32
REGION D	2764	47%	52.57	8.34	59%	2.89
REGION E	1599	74%	39.98	17.88	88%	4.23
REGION F	859	60%	29.31	1.2	63%	1.09
REGION G	333	70%	18.25	0.29	28%	0.54
REGION H	194	41%	13.93	1.1	51%	1.05
REGION I	114	40%	10.68	0.54	90%	0.73
REGION J	103	141%	10.15	0.4	27%	0.63
REGION K	564	62%	23.75	2.91	68%	1.71
REGION L	6637	73%	81.47	0.67	2%	0.82

According to the results in Table 9, regions I, H and A for paper waste give

better results with lower MAPE values than other regions. Obtained from regions MAPE values are 40%, 41% and 47%, respectively. In comparison between regions according to MAPE value, region J gives the worst results in paper waste. Regions L, J and G for plastic waste give better results with lower MAPE values than other regions. Obtained from regions MAPE values are 2%, 27% and 28%, respectively. In comparison between regions according to MAPE value, region I gives the worst results in plastic waste.

5.4. *Exponential Smoothing Method (ES)*

The exponential smoothing method is extensively utilized in the analysis of time series data (Ostertagova and Ostertag, 2012). In this thesis, the exponential smoothing method, one of the methods used to forecast time series data, is studied. The formulation of the exponential smoothing method is shown in equation 7. In this method, X_t is the actual value at time t , and F_t is the forecast value at time t . F_{t+1} is the forecast value at time $t + 1$ and the α parameter is called the smoothing constant (Ostertagova and Ostertag, 2012). The α parameter takes values between 0 and 1. The smaller the α value allows the smoother and slower changing observation to be obtained. Reducing the value of " α " can move away from the real value, while keeping it nearest to 1 may cause overfitting (Wang et al., 2023). In this study, alpha values are tried 10 times by increasing 0.1 between 0 and 1. For each region, forecasts are made by selecting the parameter that gives the best results. If α is given close to 0, the fluctuation is less visible, that is, a smoother line. On the other hand, if α is given close to 1, more fluctuations may be observed.

$$F_{t+1} = \alpha * X_t + (1 - \alpha) * F_t \quad (7)$$

In this analysis, Pandas and NumPy libraries are used for data manipulation and visualization, Statsmodels library for statistical modelling, and Matplotlib library for data visualization. For this method applied to 12 regions and two types of waste data, the graphs for just region A from both waste types are shown below. The other regions' graphics can be seen in the Appendix D. Figures 20 and 21 contain graphs according

to optimum α value results for paper and plastic waste of region A. While the optimum α value for paper waste in region A is 0.1, it is 0.4 for plastic waste.

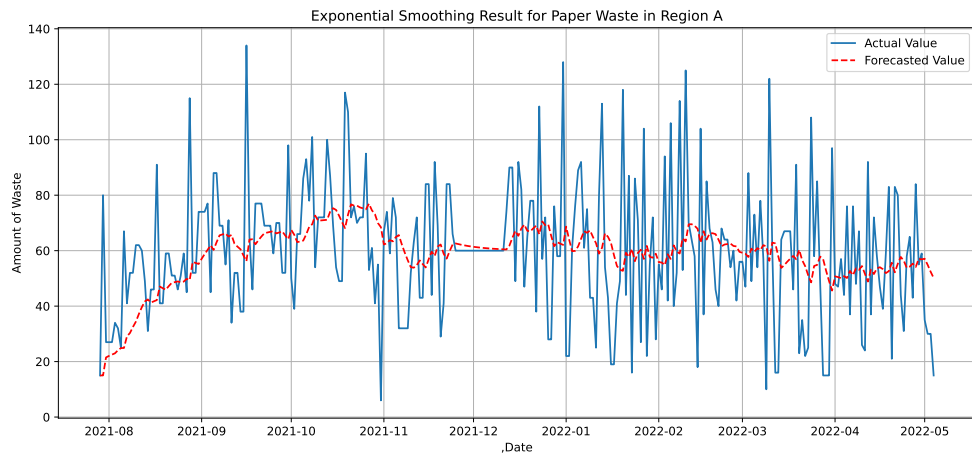


Figure 20. Exponential Smoothing Results for Paper Waste in Region A

Figure 20 shows the actual and forecasted values according to the α value of 0.1, which gives the optimum result for paper waste. The time series in blue represents the actual value. The time series in red means the forecasted value. While the optimum result is obtained when the α value is given close to 0, as can be seen from the graph, the peak points are not achieved.

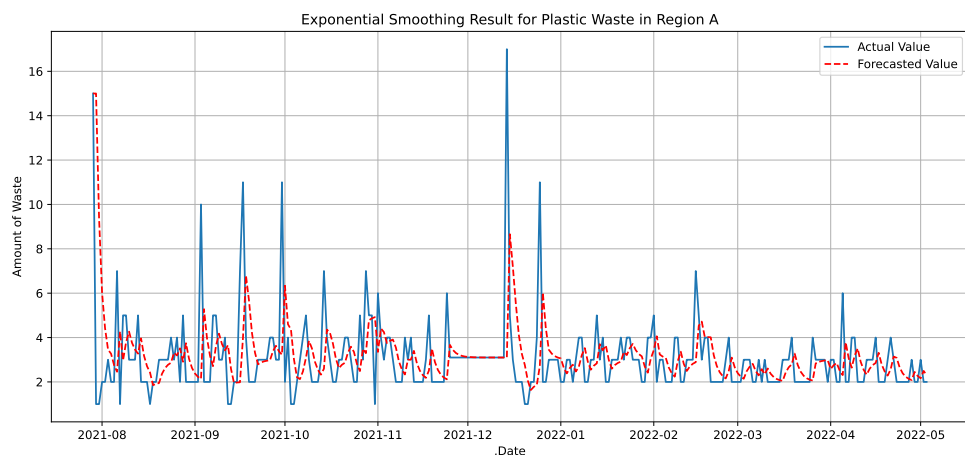


Figure 21. Exponential Smoothing Results for Plastic Waste in Region A

Figure 21 shows the actual and forecasted values according to the α value of 0.4, which gives the optimum result for plastic waste. The time series in blue represents

the actual value. The time series in red means the forecasted value. When the α value is slightly closer to 1, more peaks are reached compared to paper waste in the same region.

Table 10. Exponential Smoothing Method Results for Paper and Plastic Waste

EXPONENTIAL SMOOTHING	PAPER WASTE			PLASTIC WASTE		
	MSE	MAPE	RMSE	MSE	MAPE	RMSE
REGION A	590	45%	24.28	4.4	49%	2.10
REGION B	1126	58%	33.55	28.5	33%	5.34
REGION C	1247	59%	35.31	2.6	55%	1.61
REGION D	4168	44%	64.56	13.4	56%	3.66
REGION E	2684	74%	51.81	29.6	76%	5.44
REGION F	1422	58%	37.71	2.02	36%	1.42
REGION G	535	68%	23.13	0.49	23%	0.7
REGION H	309	40%	17.59	1.48	28%	1.23
REGION I	183	38%	13.53	0.87	62%	0.93
REGION J	163	138%	12.77	0.68	27%	0.82
REGION K	972	58%	31.18	4.6	42%	2.14
REGION L	11136	72%	105.53	1.16	6%	1.08

According to the results in Table 10, regions I, H and D for paper waste give better results with lower MAPE values than other regions. Obtained from regions MAPE values are 38%, 40% and 44%, respectively. In comparison between regions according to MAPE value, region J gives the worst results in paper waste. Regions L, G and J for plastic waste give better results with lower MAPE values than other regions. Obtained from regions MAPE values are 6%, 23% and 28%, respectively. And the region E gives the worst results in plastic waste. The best and worst results are obtained in similar regions with the MA and WMA methods.

5.5. *Holt-Winters Method (HW)*

In this section, the Holt-Winters method, one of the methods used to forecast time series data, is analyzed. This method takes into account the seasonality, trend, and irregularity (error) components of a time series and employs three parameters to forecast. These are alpha (α), beta (β), and gamma (γ). The α parameter determines how much weight is given to the current value of the time series. It takes a value

between 0 and 1, and the closer it is to 1, the greater the influence of the current observation on previous forecasts. This parameter also is used for the exponential smoothing method. The β parameter controls how smooth the trend component of the time series is. It governs the impact of the trend on previous forecasts. Similar to α , β also takes a value between 0 and 1, and the closer it is to 1, the greater the effect of the trend on current forecasts. The γ parameter regulates how smooth the seasonal component of the time series is. It influences the impact of seasonality on previous forecasts. Like α and β , γ also ranges between 0 and 1, with values closer to 1 indicating a stronger effect of seasonality on current forecasts.

Holt-Winters parameters α , β and γ are adjusted to the most optimal level by increasing 0.1 between 0 and 1 for each region and each type of waste. The seasonal period is selected 7 because the data is a time series data in daily collected. In this section, the "ExponentialSmoothing" class from the "statsmodels.tsa.holtwinters" package was used to implement the Holt-Winters method. The "sklearn.metrics" package is used to evaluate forecast performance. The "numpy" package was used for some calculations and loops. Firstly, the dataset is interpreted by separating it into time components. In Figure 22, the time component graph of the paper waste data from region A is shown and also in this figure, there are 4 subgraphs. In Figure 22, the top graph shows the behaviour of actual waste amount over time. The second subgraph involves whether there is a trend in the time series. The third subgraph observes seasonality. The fourth subgraph shows the average of the residuals (errors) collected around 0 or at the extreme points. In this section, paper and plastic waste graphs are presented only for region A. Graphs for paper and plastic waste from other regions obtained using the Holt-Winters method can be seen in Appendix E.

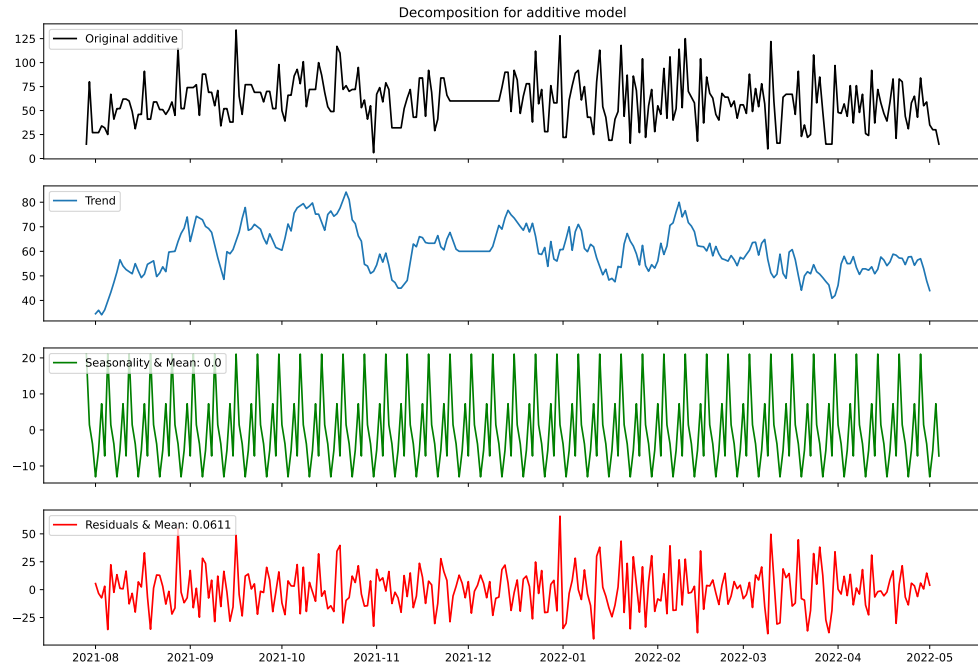


Figure 22. Decomposition of Time Series Model for Paper Waste in Region A

The statistical results are obtained for paper waste according to the graphs of time series components in region A in Figure 22. The top graph depicts the behaviour of actual waste amount over time. In the second subgraph, forecasted trend values do not represent a particular pattern through it presents peak values at certain times. When the p -value is less than 0.05, the stationarity hypothesis is rejected in the time series dataset and the time series is considered stationary. In this dataset, the p value was obtained between 0.04 and less than 0.05. It can be concluded that the time series obtained from region A for paper waste is stationary and does not contain a trend. The third subgraph shows seasonality. The fourth subgraph contains the average of the residuals (errors) collected around 0 or at the extreme points in region A.

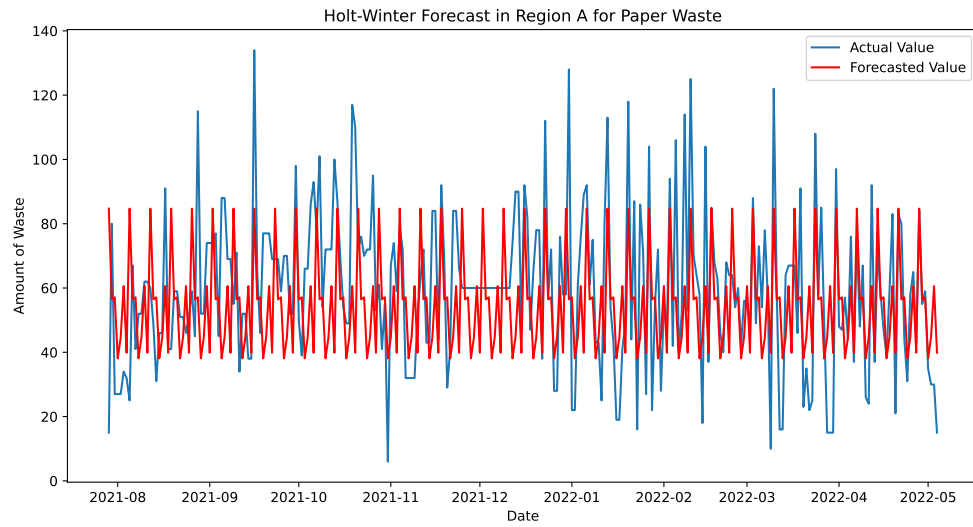


Figure 23. Holt-Winters Results for Paper Waste in Region A

To analyze weekly fluctuations throughout the month, forecasts are shown with the graph in Figure 23. For these analyses, the appropriate parameters α and β 0.0 and γ as 0.1 for paper waste in region A. According to this region analysis, since the p value is obtained lower than 0.04 and 0.05, there is no trend and it is stationary. Therefore, the beta parameter is given as 0.0. And better results are obtained when the α value is close to 0. The results showed that the MSE value is 517 and the MAPE value is 40%.

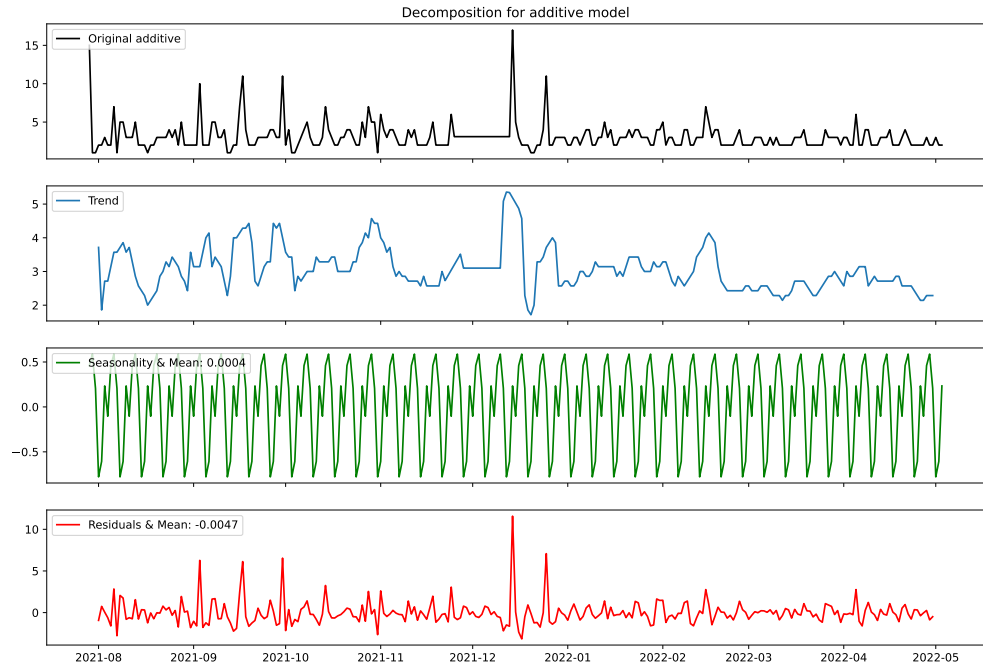


Figure 24. Decomposition of Time Series Model for Plastic Waste in Region A

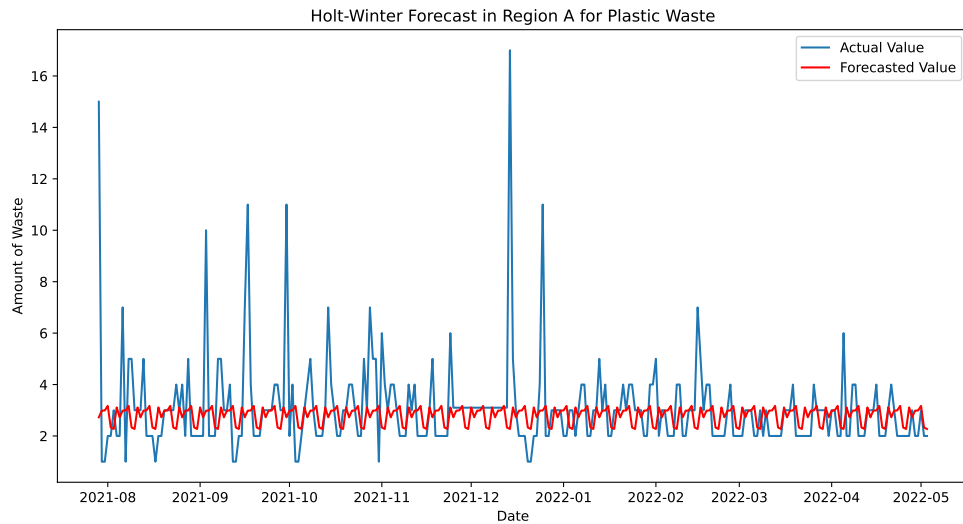


Figure 25. Holt-Winters Results for Plastic Waste in Region A

The statistical results are obtained for plastic waste according to the graphs of time series components in region A in Figure 24. The appropriate parameters α and β 0.0 and γ as 0.1 for plastic waste in Figure 25. According to this region analysis, since the

p value is obtained lower than 0.03 and 0.05, it does not contain a trend and the time series is stationary. Therefore, the beta parameter is given as 0.0. And better results are obtained when the α value is close to 0. The results showed that the MSE value is 3.6 and the MAPE value is 34%.

Table 11. Holt-Winters Results of Paper and Plastic Waste for Each Region

HOLT WINTERS	PAPER WASTE			PLASTIC WASTE		
	MSE	MAPE	RMSE	MSE	MAPE	RMSE
REGION A	517	40%	22.74	3.6	34%	1.9
REGION B	1110	56%	33.32	30.1	36%	5.5
REGION C	1008	48%	31.75	2.0	43%	1.4
REGION D	4103	44%	64.06	31	132%	5.6
REGION E	3167	116%	56.28	41	138%	6.4
REGION F	1621	79%	40.26	2.6	76%	1.6
REGION G	538	70%	23.20	0.56	32%	0.75
REGION H	288	39%	16.98	3.9	102%	1.99
REGION I	204	46%	14.27	1.1	117%	1.04
REGION J	165	139%	12.85	0.68	29%	0.08
REGION K	964	83%	31.06	15.3	126%	3.9
REGION L	13237	126%	115.05	1.2	3%	1.1

According to the results in Table 11, regions H, A and D for paper waste give better results with lower MAPE values than other regions. Obtained from regions MAPE values are 39%, 40% and 44%, respectively. In comparison between regions according to MAPE value, region J gives the worst results in paper waste. Regions L, J and G for plastic waste give better results with lower MAPE values than other regions. Obtained from regions MAPE values are 3%, 29% and 32%, respectively. And the region E gives the worst results in plastic waste. The best and worst results are obtained in similar regions with the Exponential Smoothing methods.

5.6. Comparison of Model Results Predicted by Traditional Forecast Models

Evaluated the forecast results between the traditional forecast models used and observed that the Moving Average is more successful in terms of MSE value than the other models. Tables 12 and 13 include both paper and plastic MSE, RMSE and MAPE values. According to the moving average paper waste result from the regions

where paper waste is collected, the regions that gave the lowest MAPE results were D, I and H. MAPE values were 32%, 27% and 27% for MA(3) and 39% , 33% and 24% for MA(6), respectively. In addition, weighted moving average and exponential smoothing applied to these 3 regions also obtained lower MAPE values. Holt-Winters model was obtained lower MAPE values from H, A and D regions.

Table 12. Comparison of Consistencies of Methods for Paper Waste

PAPER WASTE	MOVING AVERAGE						WEIGHTED MOVING AVERAGE			EXPONENTIAL SMOOTHING			HOLT-WINTERS		
Method Parameters	3 Days			6 Days			7 Days						Seasonality period (7 Days) (n)		
	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE
REGION A	357	32%	19	460	40%	21	385	47%	19.62	590	45%	24.28	517	40%	22.74
REGION B	615	33%	25	920	47%	30	715	60%	26.74	1126	58%	33.55	1110	56%	33.32
REGION C	765	39%	28	1.152	56%	34	856	62%	29.26	1247	59%	35.31	1008	48%	31.75
REGION D	2.532	32%	50	3.321	39%	58	2764	47%	52.57	4168	44%	64.56	4103	44%	64.06
REGION E	1.527	41%	39	2.032	59%	45	1599	74%	39.98	2684	74%	51.81	3167	116%	56.28
REGION F	810	38%	28	1.079	50%	33	859	60%	29.31	1422	58%	37.71	1621	79%	40.26
REGION G	317	51%	18	417	62%	20	333	70%	18.25	535	68%	23.13	538	70%	23.20
REGION H	172	27%	13	234	34%	15	194	41%	13.93	309	40%	17.59	288	39%	16.98
REGION I	108	27%	10	135	33%	12	114	40%	10.68	183	38%	13.53	204	46%	14.27
REGION J	97	76%	10	127	107%	11	103	141%	10.15	163	138%	12.77	165	139%	12.85
REGION K	472	35%	22	715	50%	27	564	62%	23.75	972	58%	31.18	964	83%	31.06
REGION L	5.885	38%	77	8.180	57%	90	6637	73%	81.47	11136	72%	105.53	13237	126%	115.05

Table 13. Comparison of Consistencies of Methods for Plastic Waste

PLASTIC WASTE	MOVING AVERAGE						WEIGHTED MOVING AVERAGE			EXPONENTIAL SMOOTHING			HOLT-WINTERS		
Method Parameters	3 Days			6 Days			7 Days						Seasonality period (7 Days) (n)		
	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE
REGION A	1.9	30%	1.4	2.6	36%	1.6	2.39	45%	1.55	4.4	49%	2.10	3.6	34%	1.9
REGION B	16	22%	4	21.1	27%	4.6	17.08	33%	4.13	28.5	33%	5.34	30.1	36%	5.5
REGION C	1.6	38%	1.3	2.3	50%	1.5	1.73	56%	1.32	2.6	55%	1.61	2.0	43%	1.4
REGION D	7.7	33%	2.8	10.8	47%	3.3	8.34	59%	2.89	13.4	56%	3.66	31	132%	5.6
REGION E	16.1	43%	4	20.6	65%	4.5	17.88	88%	4.23	29.6	76%	5.44	41	138%	6.4
REGION F	0.99	23%	1	1.7	44%	1.3	1.2	63%	1.09	2.02	36%	1.42	2.6	76%	1.6
REGION G	0.25	13%	0.5	0.38	20%	0.6	0.29	28%	0.54	0.49	23%	0.7	0.56	32%	0.75
REGION H	0.78	22%	0.9	1.48	36%	1.2	1.1	51%	1.05	1.48	28%	1.23	3.9	102%	1.99
REGION I	0.4	41%	0.6	0.69	68%	0.8	0.54	90%	0.73	0.87	62%	0.93	1.1	117%	1.04
REGION J	0.39	15%	0.6	0.55	24%	0.7	0.4	27%	0.63	0.68	27%	0.82	0.68	29%	0.08
REGION K	2.02	32%	1.4	3.47	50%	1.9	2.91	68%	1.71	4.6	42%	2.14	15.3	126%	3.9
REGION L	0.13	2%	0.4	0.03	1%	0.2	0.67	2%	0.82	1.16	6%	1.08	1.2	3%	1.1

CHAPTER 6: FORECASTING WITH MACHINE LEARNING

In this chapter, forecasts were performed using XGBoost, K-NN and ARIMA methods, which are machine learning algorithms, based on the collection amounts obtained from the dataset of 12 regions and 2 types of data. The results of the algorithms are compared with the RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error) and MSE (Mean Squared Error) values defined in Section 5.1

6.1. *Supervised Learning and Selected Methods for Analysis*

Machine learning is mainly divided into subcategories supervised learning, unsupervised learning and reinforcement learning. Supervised learning focuses on training with labelled data, whereas unsupervised learning deals with learning from unlabeled data without predefined output (Son et al., 2023). Reinforcement learning, a type of machine learning, trains agents to make decisions within an environment based on the results of their actions (Klein et al., 2024). Due to the feature of the dataset used in this thesis, the concept of supervised learning will be emphasized. Supervised learning involves teaching machines by providing appropriately labelled training data and utilizing this data to forecast the outcome (Das et al., 2024). In regression problems, there are numerical values y (output value) that we will predict. Numerical forecasts are made using x (input value) values. In classification problems, the y class is tried to be predicted categorically by using x values. For example, when predicting the weather in regression, the result is a numerical degree, while in classification, an answer such as hot or cold is given. There are the expected output values of each observation along with the input features in the training dataset. In the waste dataset, for example, each waste sample has a target variable (output value) that contains the actual amount of waste. The features of the data predicted by machine learning can be listed as the month, day, year and day of the week when the waste was collected. Analysis of the 'Month' feature in the dataset shows the effects of seasonal changes on waste quantities. The "day of the week" feature highlights trends in waste production

observed on certain days of the week.

XGBoost is a machine learning algorithm that is effective in both regression and classification problems in the literature. It provides strong performance and is therefore a preferred method in many areas. XGBoost can identify complex relationships within data and build a powerful model to predict these relationships (Hong et al., 2023). It is known for fast training times and generally highly accurate predictions (Zhang et al., 2022). It is preferred in this study to obtain fast and accurate waste amount forecast. K-Nearest Neighbour (K-NN) is a machine learning algorithm that provides simplicity and high flexibility. The algorithm is introduced in the literature as one of the top 10 algorithms in data mining (Xu et al., 2019). Since it is a data-based learning approach, it is generally preferred for simple data structures. In estimating waste amounts, it can be considered that waste amounts with similar characteristics are close to each other. K-NN can make forecasting by taking these similarities into account. ARIMA is a widely used model in time series analysis. It is effective for predicting trends, seasonality, and other regularities of time series Punyapornwithaya et al. (2023). Waste amounts are generally data that changes over time. The waste dataset used in the study consists of discrete time series data. ARIMA, a time series model, is a suitable option for forecasting such data. As a result, since waste quantity data is time series type data, it can be analyzed using time series analysis methods. ARIMA (Autoregressive Integrated Moving Average) is a method used in time series forecasting in machine learning. K-NN (K-Nearest Neighbors) and XGBoost are used in both classification and regression problems in the literature. These three methods are studied in this thesis.

6.1.1. *Extreme Gradient Boosting (XGBoost)*

Gradient Boost Decision Tree (GBDT) is the foundation of XGBoost, a highly scalable machine learning method that was first introduced by Chen and Guestrin (2016). XGBoost, introduced in 2016, is regarded as one of the best algorithms for supervised learning involving classification and regression (Wan et al., 2024). In the literature, prediction studies have been conducted on various situations using XGBoost (Zhang et al., 2022; Hong et al., 2023). XGBoost objective function is described as

follows (Hong et al., 2023):

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (1)$$

Equation 1, l represents a differentiable convex loss function, quantifying the difference between the forecast $\hat{y}_i$ and the target y_i ; Ω signifies the regularization term integrated into the objective function to manage the model's complexity and prevent overfitting; and f_k denotes the function of the k -th tree. f_k is the forecast function of a given tree. XGBoost uses a gradient-boosting algorithm by creating multiple decision trees. Each tree is applied to a subset of the dataset and forecasts the residuals from the predictions of previous trees. Each tree is often optimally tuned to reduce debris from previous trees. Therefore, f_k represents the forecast function of a particular tree.

6.1.2. *K-Nearest Neighbors (K-NN)*

K-Nearest Neighbors (K-NN) is an algorithm used in regression and classification problems in supervised learning machine learning models that makes predictions based on the target output of the nearest neighbours of a given query point (Twumasi and Twumasi, 2022). The stages of the model are as firstly the number of neighbours (k) is determined. Secondly, the distance between the unknown point and all other points is calculated. Afterwards, the distances are arranged in ascending order, and the k nearest observations are chosen. Lastly, classification is most often given as the class and regression is given as the mean value forecasting value. To explain the model objective function, assume the i -th training instance or data point involves a vector of v features: $m = (m_1^i, m_2^i \dots m_v^i)$ which explains the example. Given a new example with known features: $l = (l_1, l_2 \dots, l_v)$ however unknown the target vector. These new features are used to obtain its most similar training examples from their vectors of features and the Euclidean distance as in equation 2 (Twumasi and Twumasi, 2022). Among the distance calculation metrics types, there are used in general such as Euclidean distance, Manhattan Distance and Minkowski distance. Euclidean distance was applied in this thesis.

$$d(m, l) = \sqrt{\sum_{j=1}^v (m_j^i - l_j)^2} \quad (2)$$

The nearest training data points are chosen, and a prediction is generated by averaging the target output values associated with these points. That is the targets $(y_1, y_2, \dots, y_K)$ are the K nearest neighbours for a new example x , and then the forecast $(\hat{y})$ is given by in equation 3 (Twumasi and Twumasi, 2022).

$$\hat{y} = \frac{1}{K} \sum_{j=1}^K y_i \quad (3)$$

6.1.3. Autoregressive Integrated Moving Average (ARIMA)

The ARIMA model, another time series analysis, is used to forecast the amounts of the waste dataset. ARIMA model has autoregressive (AR), integrated (I) and moving average (MA) components and includes p , d and q hyperparameters (Punyapornwithaya et al., 2023). p (AR - AutoRegressive) hyperparameter specifies the number of autoregressive terms in the model. Autoregressive terms represent the association of dependent variable values at past time steps with the current value. In other words, the p -value determines how many steps back are considered to predict the current value. d (I - Integrated) hyperparameter specifies the degree of differencing terms in the model. Differencing terms are used to ensure whether the data is stationary or to remove the trend in the model. The d value determines how much differencing is required to make the time series stationary. q (MA - Moving Average) hyperparameter specifies the number of moving average terms in the model. Moving average terms represent the association of past error terms with the current error term. Hence, the q value determines how many steps back are considered to predict the current error term. Forecasting is made by dividing the dataset into test and training sets with the parameters p , d and q suitable for the model. The build of the ARIMA process is written in equation 4 (Punyapornwithaya et al., 2023):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \Theta_2 \varepsilon_{t-2} - \dots - \Theta_q \varepsilon_{t-q} \quad (4)$$

In equation 4, y_t is the observation at time t , ϕ and Θ are the parameters of AR and MA, respectively, p and q and ε is the residual series. ε represents the error terms in the ARIMA model. These terms express how much the model's forecasts deviate from the actual values. While the ARIMA model attempts to model the structural features and trends of time series data, error terms express the uncertainties of the model. These errors are used to evaluate how well the model fits the real data and how accurate the predictions are.

6.2. Validation of Models

In this thesis, when the dataset is analyzed, the dependent variable is the amount of waste quantity. The collection day, month and year are specified as independent variables. Firstly, the waste amounts as the dependent variable of each region regarding the waste data and the selected independent variables were taken into consideration in the model to be trained. Waste amounts increase or decrease depending on time in each region. Then the dataset was divided into training data and testing data. Training data was applied to build models while testing data was used to verify the validity of the built models.

In the literature, K -fold cross-validation is used to verify the performance of the machine learning model (Qambar and Al Khalidy, 2023). The dataset is subjected to split into K subsets of similar sizes. In this way, different parts of the dataset are selected in each k -fold iteration and the model is trained on different data sections. Training the model only on a certain part of the dataset may cause overfitting (Wang et al., 2023). In this thesis, the k -fold cross-validation method is used to avoid overfitting. Typically, the K value is set at five or ten (Marcot and Hanea, 2021). The K value was chosen as five for this study. In XGBoost and K-NN models, 5-fold cross-validation to the waste dataset is applied to reduce overfitting in the forecasting study of waste amount. For the ARIMA model, the hold-out method is used. The hold-out method is a widely use cross-validation technique to evaluate the performance of a time series model. In this method, the dataset is usually divided into training and test sets. For time series models such as ARIMA, the hold-out method is commonly used to evaluate the accuracy of the model. While historical data is used to train the model,

the test set usually consists of recent data. The model's forecasts are evaluated on the test set and the performance of the model is determined by measuring how compatible the forecasts are with the real values.

In this study, the pandas package in Python is used for data manipulation and analysis. The "KFold" class is used for k -fold cross-validation. The "cross_val_score" function is used to evaluate the performance of a model by cross-validation. The packages used include XGBRegressor, KNeighborsRegressor, and ARIMA for training models. The "matplotlib.pyplot" module is used to visualize data.

6.3. Comparison of Model Results Forecasted by Machine Learning

6.3.1. Extreme Gradient Boosting (XGBoost) Results

Two types of waste data amounts for 12 regions are forecasted with the XGBoost method. In this thesis, default parameters are utilized in "XGBRegressor" class of XGBoost. First, the paper data forecast results will be interpreted, and then the plastic waste data results will be given. The dataset was trained using the 5-fold cross-validation method and train results are shown with MAPE in Table 14 and 17. With 5-fold cross-validation, the dataset for 12 regions was trained and 5 different MSE values were obtained according to the forecast results on the test set. The average of the MSE values obtained for paper and plastic waste for each region are shown in tables 15 and 18. According to the average MSE value obtained from the test set, forecasting results are found for each region. The actual and forecasted values were compared on the test set.

In the XGBoost model, train successes shown in Tables 14 and 17 are quite different compared to test successes. For example, in region A for paper waste, while the train result was 0.50%, the test result was 50.39%. The use of default parameters in this model may be a factor in the difference between train and test results. The default parameters of algorithms such as XGBoost are generally suitable for general datasets, but the best parameters for each dataset and problem may be different. Default parameters can often cause the model to overfit a particular dataset. If the default parameters cause the model to fit the training data too well, its performance on the

test data may be poor. This may result in a result similar to overfitting. Therefore, to improve the performance of the model, it may be necessary to carefully adjust the default parameters or try different parameter values. By using hyperparameter optimization techniques such as grid search or random search, it is possible to find the best parameter values. In this way, the difference between the train and test results of the model can be reduced and its overall performance can be improved.

Table 14. Train Set Results of Paper Waste using XGBoost

Region Name	Train Set Results (MAPE Values)
Region [A]	0.48%
Region [B]	0.27%
Region [C]	0.37%
Region [D]	0.46%
Region [E]	0.51%
Region [F]	0.75%
Region [G]	0.78%
Region [H]	0.32%
Region [I]	0.26%
Region [J]	0.57%
Region [K]	0.41%
Region [L]	0.60%

The average MSE results obtained with 5-fold cross-validation from paper waste are shown in the table below.

Table 15. Average MSE Values Obtained From 5-Fold Cross-Validation for Paper Waste

Region Name	Average MSE Values in Paper Waste
Region [A]	666
Region [B]	1496
Region [C]	1264
Region [D]	5123
Region [E]	5572
Region [F]	2796
Region [G]	622
Region [H]	450
Region [I]	247
Region [J]	243
Region [K]	1586
Region [L]	22015

Graphs showing the comparison of actual and forecasted results for paper waste are presented in Figure 26, 27 and 28. These graphs belong to the 3 regions where the best results were obtained, the other regions' graphics are in Appendix F of the thesis. When examined by region, the three regions that give the best results are "Region H", "Region C" and "Region A", respectively. MAPE values were 47.41%, 48.10% and 50.39%, respectively.

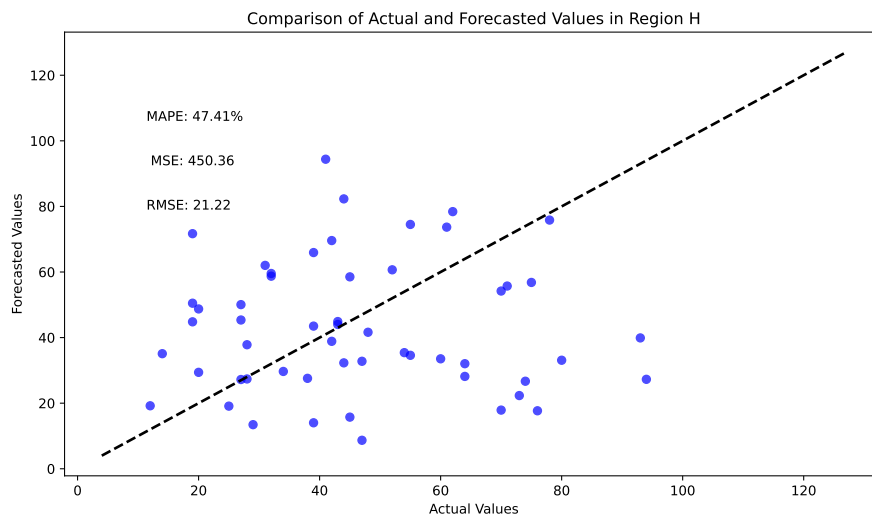


Figure 26. Graphs of Actual and Forecasted Values in the Testset of Region H

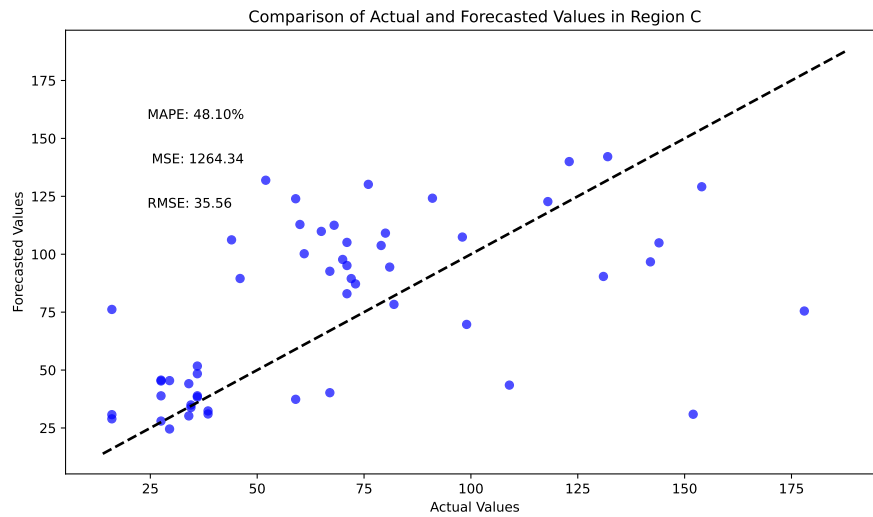


Figure 27. Graphs of Actual and Forecasted Values in the Testset of Region C

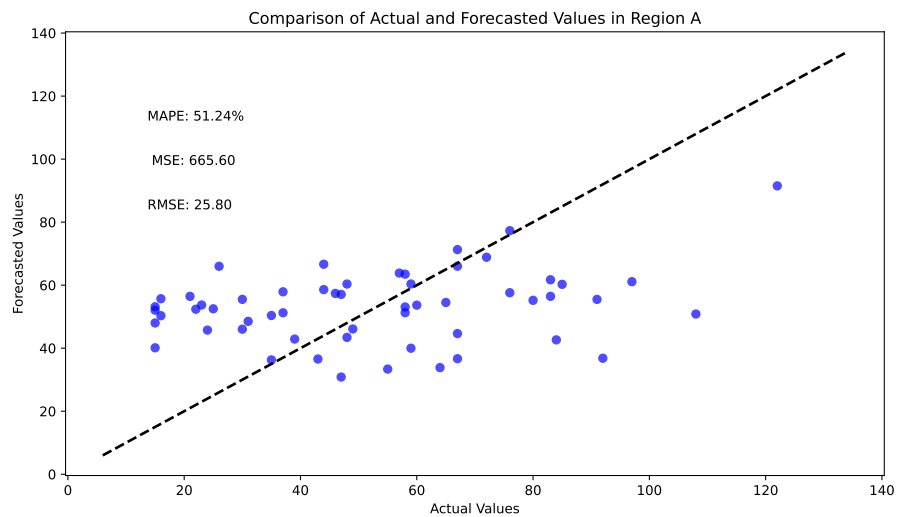


Figure 28. Graphs of Actual and Forecasted Values in the Testset of Region A

Table 16 shows paper waste forecast results with MSE, MAPE and RMSE values obtained from XGBoost. According to the forecast results in the test set, the best result among the regions is the H region with a MAPE value of 47.41%.

Table 16. Paper Waste Forecasting Results using XGBoost

Region Name	MSE	MAPE	RMSE
Region [A]	666	51.24%	25.80
Region [B]	1496	66.25%	38.68
Region [C]	1264	48.10%	35.56
Region [D]	5123	50.56%	71.58
Region [E]	5572	135.57 %	74.65
Region [F]	2796	109.25 %	52.88
Region [G]	622	67.06 %	24.93
Region [H]	450	47.41 %	21.22
Region [I]	247	53.10 %	15.73
Region [J]	243	144.31 %	15.59
Region [K]	1586	73.46 %	39.83
Region [L]	22015	148.05 %	148.37

The dataset for plastic waste was trained using the 5-fold cross-validation method and train results are shown with MAPE in Table 17. Similar to paper results, There is a huge difference between train and test results. For example in region A for plastic waste, while the training result was 0.38%, the test result was 51.36%.

Table 17. Train Set Results for Plastic Waste using XGBoost

Region Name	Train Set Results (MAPE Values)
Region [A]	0.38%
Region [B]	0.31%
Region [C]	0.40%
Region [D]	0.41%
Region [E]	0.16%
Region [F]	0.13%
Region [G]	0.28%
Region [H]	0.23%
Region [I]	0.44%
Region [J]	0.17%
Region [K]	0.28%
Region [L]	0.01%

The average MSE results obtained with 5-fold cross-validation from plastic waste

are shown in the table below.

Table 18. Average MSE Values Obtain From 5-Fold Cross Validation for Plastic Waste

Region Name	Average MSE Values in Plastic Waste
Region [A]	4.85
Region [B]	49.32
Region [C]	2.99
Region [D]	19.65
Region [E]	121.42
Region [F]	3.56
Region [G]	0.83
Region [H]	7.24
Region [I]	1.58
Region [J]	1.21
Region [K]	10.94
Region [L]	34.05

Graphs showing the comparison of actual and forecasted results for plastic waste are presented in Figure 29, 30 and 31. These graphs belong to the 3 regions where the best results were obtained, the other regions' graphics are in Appendix F of the thesis. When examined by region, the three regions that give the best results are "Region G", "Region J" and "Region B", respectively. MAPE values were 36.82%, 42.69% and 46.60%, respectively.

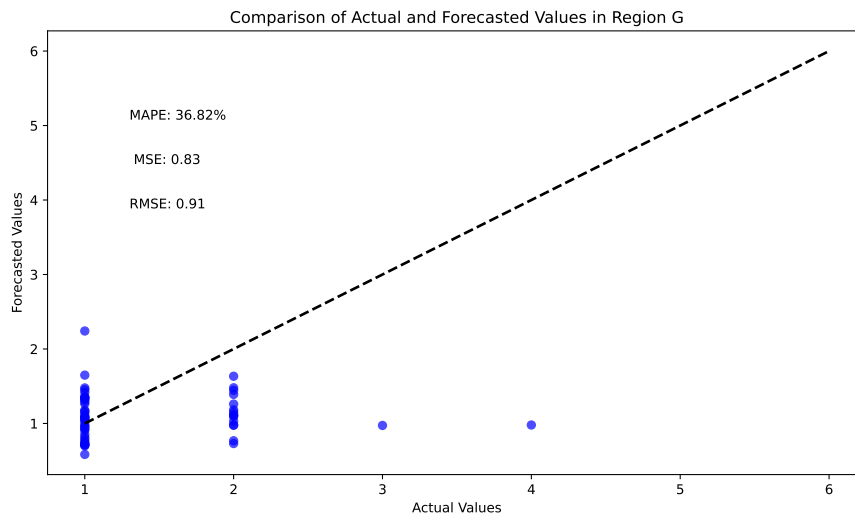


Figure 29. Graphs of Actual and Forecasted Values in the Testset of Region G

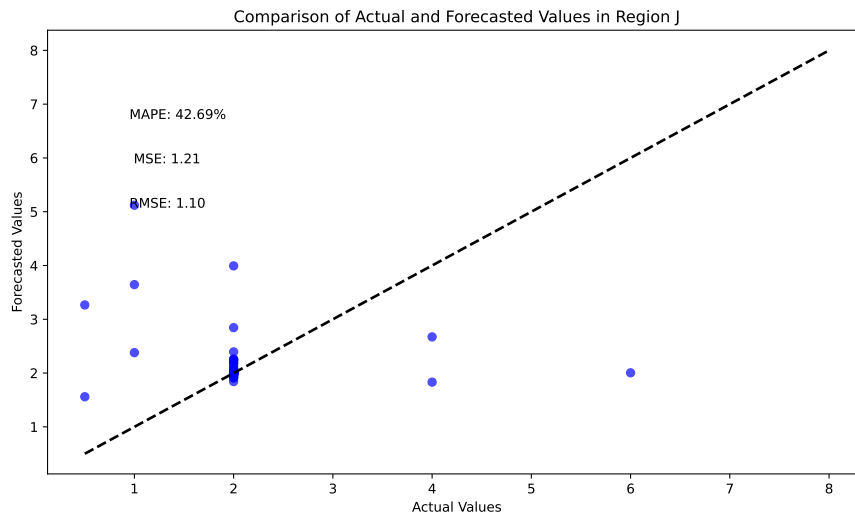


Figure 30. Graphs of Actual and Forecasted Values in the Test Set of Region J

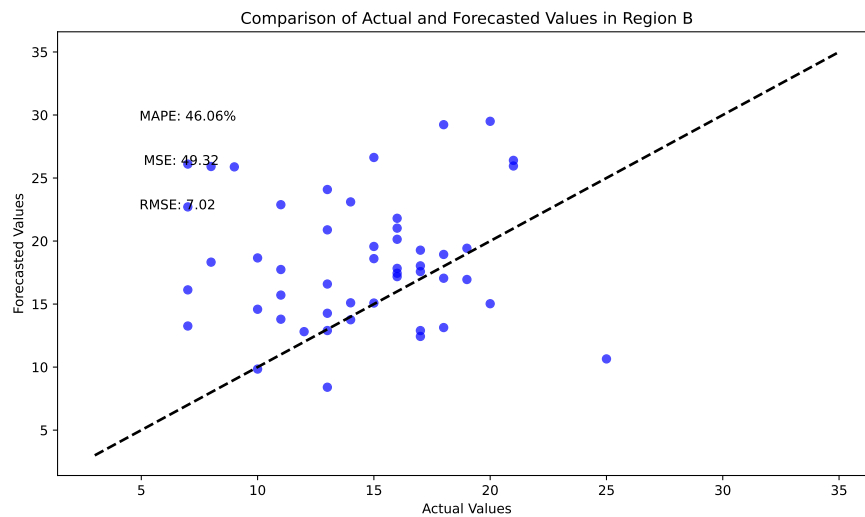


Figure 31. Graphs of Actual and Forecasted Values in the Testset of Region B

Table 19 shows plastic waste forecast results with MSE, MAPE and RMSE values obtained from XGBoost. According to the forecast results in the test set, the best result among the regions is the G region with a MAPE value of 36.82%.

Table 19. Plastic Waste Forecasting Results using XGBoost

Region Name	MSE	MAPE	RMSE
Region [A]	4.85	51.36%	2.20
Region [B]	49.32	46.60%	7.02
Region [C]	2.99	48.11%	1.73
Region [D]	19.65	60.95%	4.43
Region [E]	121.42	263.32%	11.02
Region [F]	3.56	73.67%	1.89
Region [G]	0.83	36.82 %	0.91
Region [H]	7.24	132.41 %	2.69
Region [I]	1.58	118.21%	1.26
Region [J]	1.21	42.69 %	1.10
Region [K]	10.94	108.21 %	3.31
Region [L]	34.05	163.89%	5.84

6.3.2. *K-Nearest Neighbors (K-NN) Results*

The waste data amounts belonging to two types for 12 regions are forecasted with the K-NN method. The model performance is evaluated for different k values using K -fold cross-validation to choose the optimal k value. According to a review of the K-NN method conducted by Chirici et al. (2016), the most frequently used number of neighbours in remote sensing applications is $k=5$, although values ranging from 1 to 10 are commonly observed. Additionally, a hybrid k -nearest neighbour (H-KNN) approach was proposed by Paulauskaite-Taraseviciene et al. (2022) to estimate municipal solid waste, and k values were tested from 1 to 9. Taking the mentioned articles as an example in this thesis in the K-NN model, the k parameter was tested from 1 to 10. It is seen that as the K value increases, the MSE value decreases. However, as increasing the K value may cause overfitting, a larger value was not selected (Wang et al., 2023). In addition, the grid search is employed to optimize parameters and enhance the model's efficiency (Belete and Huchaiah, 2022). It's a prevalent approach used to explore the hyperparameter configuration space. In this Section, the Grid search technique is used to make forecasts by finding the optimal k value between 1 to 10 for each region and data type.

First of all, the dataset of 12 regions is trained with 5-fold cross-validation with optimized parameters. Train results were analyzed according to 5-fold cross-validation and the average of 5 different results from each region is shown in the table 20. According to Train results, the regions that gave the best results were H, D and I regions. MAPE values were 34.32%, 39.65% and 42.92%, respectively.

The forecast is made using the best k value selected for each region according to the grid search technique, and the 5-fold average MSE values are shown in Tables 21 and 24. The graphs according to the forecasting results for paper waste with the average MSE values obtained from 5-Fold cross-validation are presented in Figure 32, 33 and 34. These graphs belong to the 3 regions where the best results were obtained, the other regions' graphics are in Appendix G of the thesis.

Table 20. Train Set Results of Paper Waste using K-NN

Region Name	Train Set Results (MAPE Values)
Region [A]	44.99%
Region [B]	47.64%
Region [C]	53.70%
Region [D]	39.65%
Region [E]	86.44%
Region [F]	68.88%
Region [G]	60.08%
Region [H]	34.32%
Region [I]	42.92%
Region [J]	121.67%
Region [K]	47.20%
Region [L]	97.84%

Table 21. Average MSE Values Obtain From 5-Fold Cross-Validation for Paper Waste

Region Name	Average MSE Values in Paper Waste
Region [A]	577
Region [B]	1169
Region [C]	1394
Region [D]	4414
Region [E]	3879
Region [F]	2013
Region [G]	672
Region [H]	395
Region [I]	236
Region [J]	182
Region [K]	999
Region [L]	17824

In Figure 32, 33 and 34, points that coincide with the line indicate that the forecasted values are equal to the actual values. Points close to the line indicate that the forecast results are close to actual values, while points that are farther away indicate that the difference between the forecasted value compared to the actual value is larger than the others.

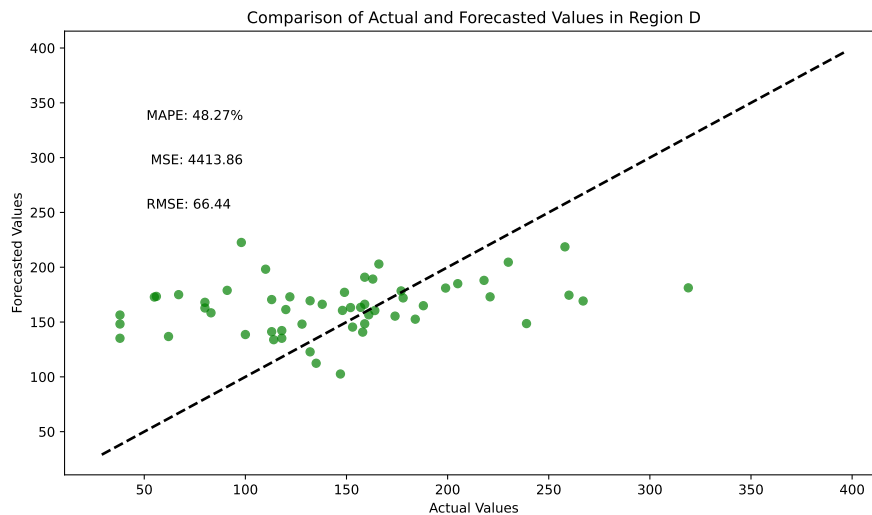


Figure 32. Graph of Forecasted Data Corresponding to Actual Data of Paper Waste in Region D

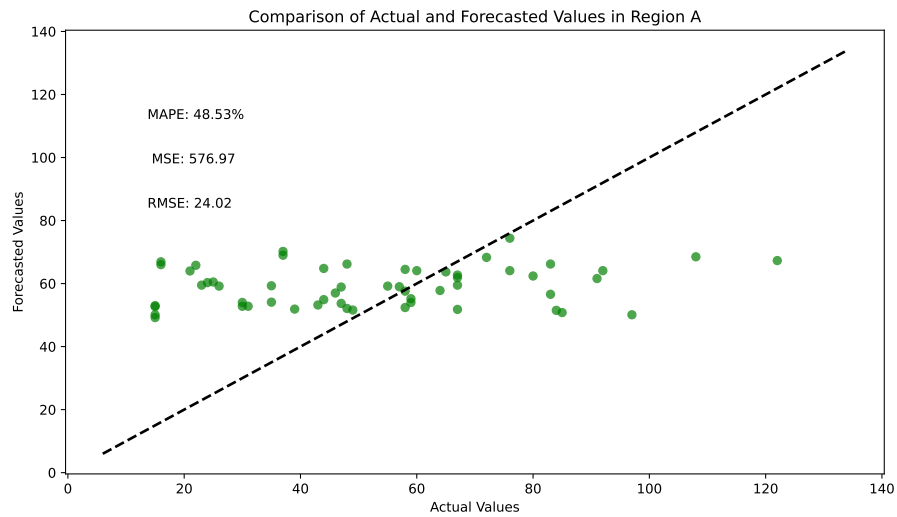


Figure 33. Graph of Forecasted Data Corresponding to Actual Data of Paper Waste in Region A

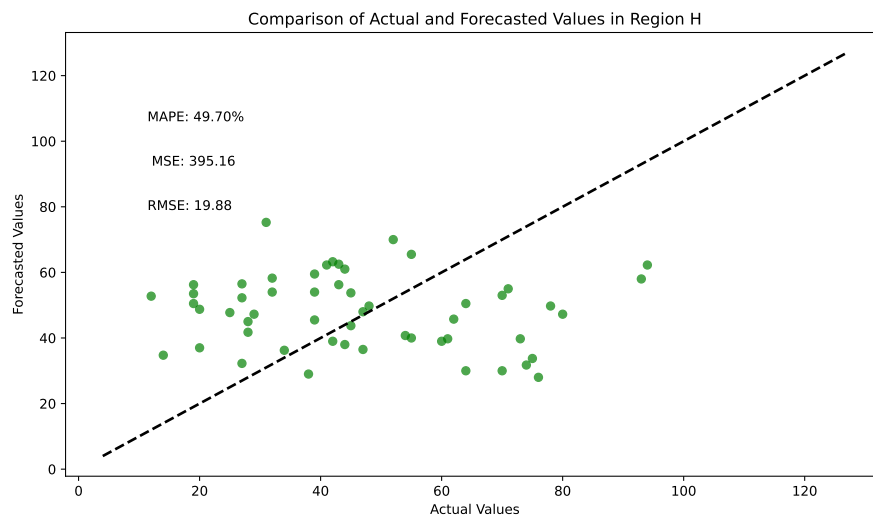


Figure 34. Graph of Forecasted Data Corresponding to Actual Data of Paper Waste in Region H

Table 22 shows paper waste forecast results with MSE, MAPE and MSE values obtained from K-NN. When examined by region, the 3 regions with the best results are "Region D", "Region A", and "Region H", respectively. MAPE values were 48.27%, 48.68% and 49.70 %, respectively. Values close to the train results were obtained.

Table 22. Paper Waste Forecasting Results Obtained from Testset using K-NN

Region Name	MSE	MAPE	RMSE
Region [A]	577	48.53%	24.02
Region [B]	1169	62.55%	34.19
Region [C]	1394	66.30%	37.34
Region [D]	4414	48.27%	66.44
Region [E]	3879	119.40 %	62.28
Region [F]	2013	89.89 %	44.87
Region [G]	672	74.65 %	25.92
Region [H]	395	49.70 %	19.88
Region [I]	236	51.14 %	15.37
Region [J]	182	135.66 %	13.50
Region [K]	999	78.67 %	31.60
Region [L]	17824	141.33 %	133.51

Train results for plastic waste were analysed according to 5-fold cross-validation and the average of 5 different results from each region is shown in the table 23.

Table 23. Trainset Results of Plastic Waste using K-NN

Region Name	Train Set Results (MAPE Values)
Region [A]	39.49%
Region [B]	32.64%
Region [C]	50.10%
Region [D]	56.21%
Region [E]	116.81%
Region [F]	53.21%
Region [G]	28.70%
Region [H]	79.18%
Region [I]	79.09%
Region [J]	26.78%
Region [K]	71.50%
Region [L]	6.79%

Table 24. Average MSE Values Obtain From 5-Fold Cross-Validation for Plastic Waste

Region Name	Average MSE Values in Plastic Waste
Region [A]	3.69
Region [B]	40.24
Region [C]	2.88
Region [D]	15.61
Region [E]	48.37
Region [F]	3.63
Region [G]	0.70
Region [H]	6.04
Region [I]	1.22
Region [J]	0.73
Region [K]	7.11
Region [L]	1.39

Graphs showing the comparison of actual and forecasted results for plastic waste are presented in Figure 35, 36 and 37. These graphs belong to the 3 regions where the best results were obtained, the other regions' graphics are in Appendix G of the thesis. When examined by region, the three regions that give the best results are "Region L", "Region J" and "Region B", respectively. MAPE values were 14.32%, 33.32% and 42.62%, respectively.

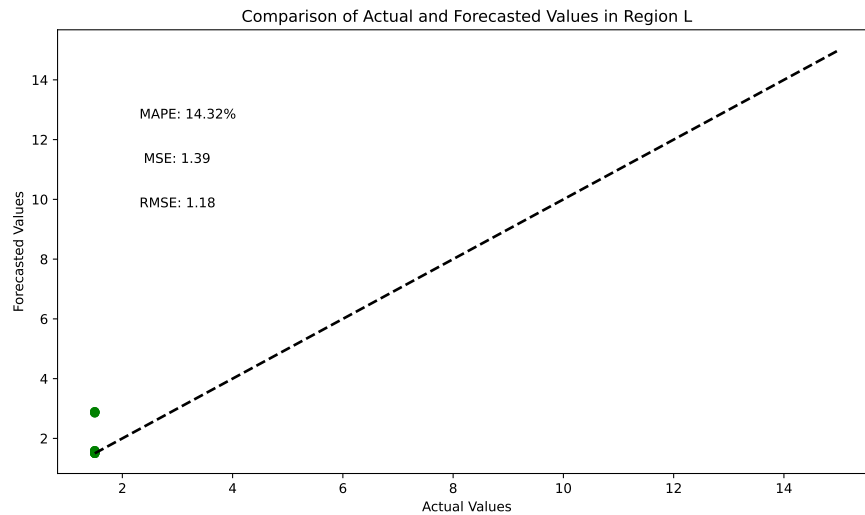


Figure 35. Graph of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region L

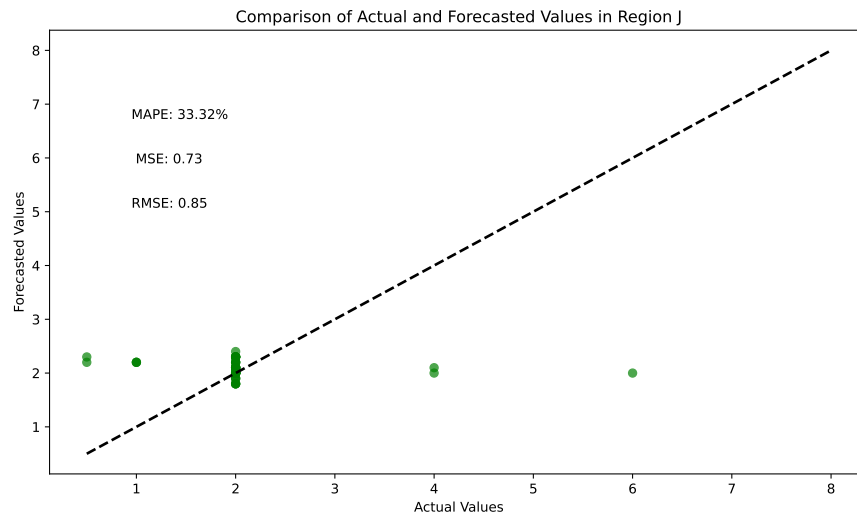


Figure 36. Graph of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region J

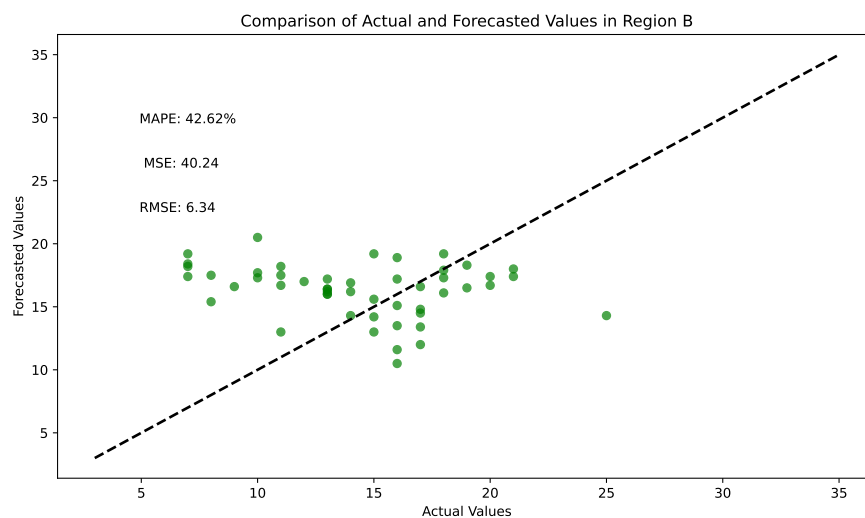


Figure 37. Graph of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region B

Table 25. Plastic Waste Forecasting Results Obtained from Testset using K-NN

Region Name	MSE	MAPE	RMSE
Region [A]	3.69	46.55%	1.92
Region [B]	40.24	42.62%	6.34
Region [C]	2.88	58.57%	1.70
Region [D]	15.61	70.25%	3.95
Region [E]	48.37	179.19%	6.95
Region [F]	3.63	84.22%	1.91
Region [G]	0.70	40.39%	0.84
Region [H]	6.04	127.92%	2.46
Region [I]	1.22	104.30%	1.11
Region [J]	0.73	33.32%	0.85
Region [K]	7.11	93.99%	2.67
Region [L]	1.39	14.32%	1.18

6.3.3. Autoregressive Integrated Moving Average (ARIMA) Results

In this thesis, the parameter selection for the ARIMA model was conducted automatically using the "*auto – arima*" function from the "pmdarima" library. This function applies a model selection algorithm to automatically choose the most suitable ARIMA parameters based on time series data. The ARIMA model forecasts future values based on past values and error terms of a time series. Therefore, it is important to protect the internal structure of the time series and allow predictions to be made based on historical data. While making predictions, the dataset was set as 80% train and 20% test set. The results of the Train set are shown in the table 26 and 28 for each region. According to train results, the best results in paper waste are obtained from the H region with 34.45%. Plastic waste is obtained from the L region with 7.84%.

There are graphs according to the forecasting results for paper waste obtained from ARIMA in Figure 38, 39 and 40. These graphs belong to the 3 regions where the best results were obtained, the other regions' graphics are in Appendix H of the thesis. When examined by region, the 3 regions with the best results are "Region L", "Region I", and "Region J ", respectively. MAPE values were 29.99%, 31.39% and 42.69% respectively. The model was run by selecting parameters (2,1,1) for Region L,

parameters (1,0,1) for Region I and parameters (0,0,1) for Region J. If an interpretation of the results according to the regions, test results were obtained close to the train results.

Table 26. Trainset Results of Paper Waste using ARIMA

Region Name	Train Set Results (MAPE Values)
Region [A]	41.89%
Region [B]	54.43%
Region [C]	40.63%
Region [D]	42.32%
Region [E]	79.96%
Region [F]	58.01%
Region [G]	68.18%
Region [H]	34.45%
Region [I]	41.99%
Region [J]	145.85%
Region [K]	65.70%
Region [L]	78.48%

In Figures 38, 39 and 40 the forecast results on the test set are compared with the actual values. While the blue line in the graphs means actual value, the yellow line indicates forecasted values in time-dependent.

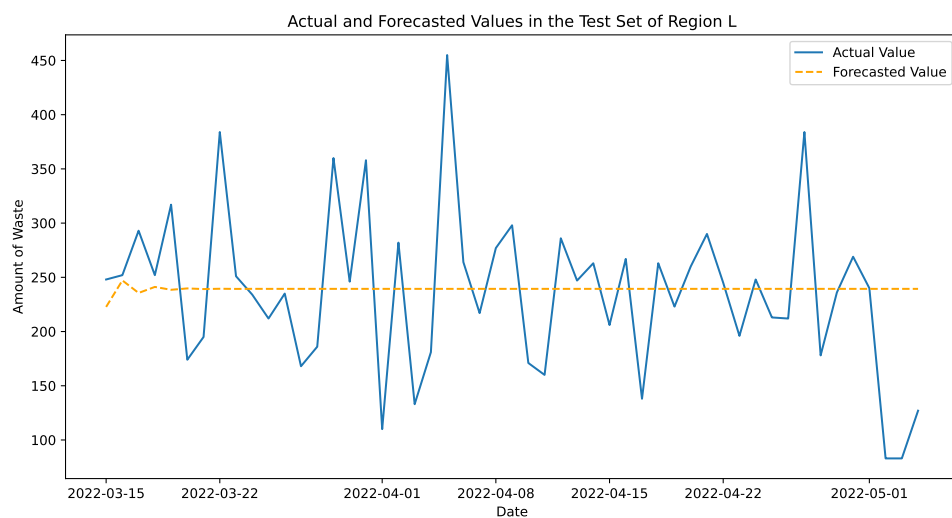


Figure 38. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region L

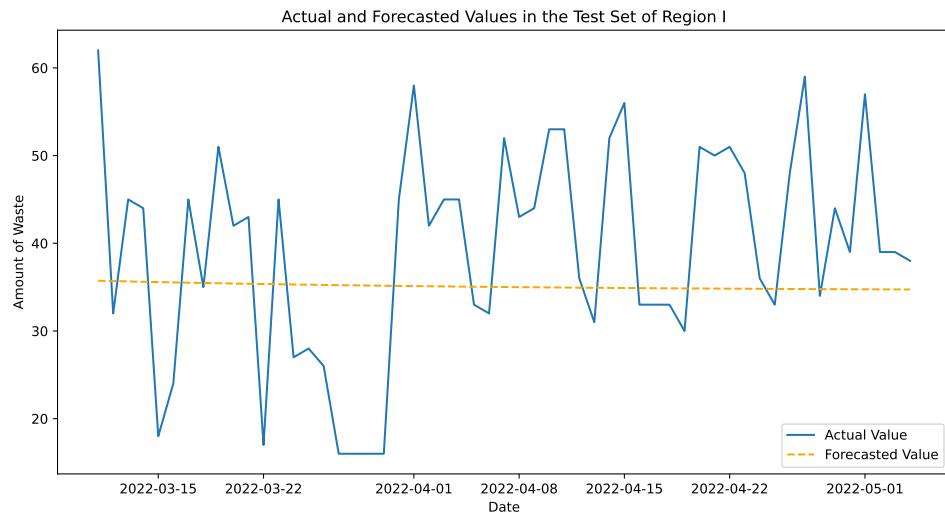


Figure 39. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region I

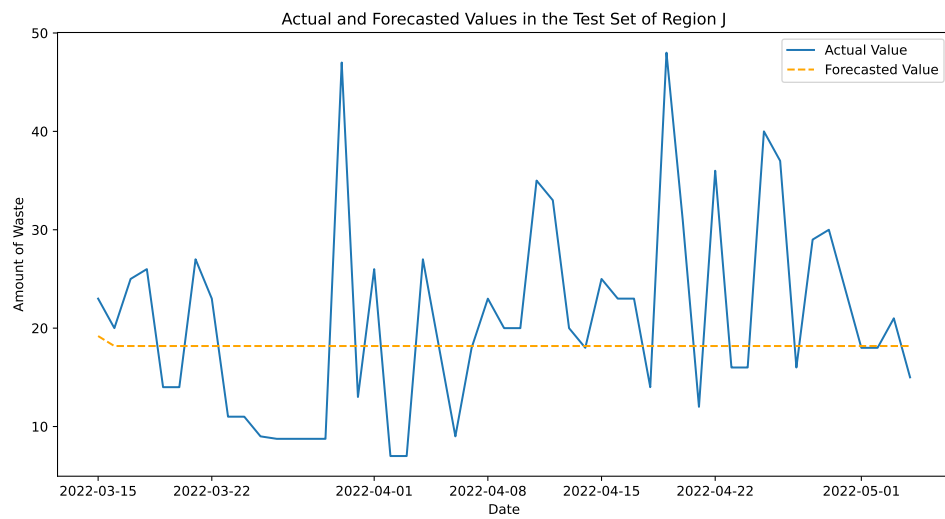


Figure 40. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region J

Table 27. Paper Waste Forecasting Results Obtained from Testset using ARIMA

Region Name	MSE	MAPE	RMSE
Region [A]	725	73.26%	26.9
Region [B]	1139	50.33%	33.7
Region [C]	1369	70.61%	37
Region [D]	3946	52.84%	62.8
Region [E]	2395	66.41%	48.9
Region [F]	2341	66.87%	48.3
Region [G]	773	68.15%	27.8
Region [H]	417	46.89%	20.4
Region [I]	163	31.39%	12.8
Region [J]	105	42.69%	10.2
Region [K]	669	63.90%	25.9
Region [L]	5509	29.99%	74.2

In Figure 28 there are train results regarding plastic waste. There was a big difference between the test results compared to the train results of some regions. This may be because plastic waste data is less than paper waste data.

Table 28. Trainset Results for Plastic Waste using ARIMA

Region Name	Train Set Results (MAPE Values)
Region [A]	52.11%
Region [B]	34.49%
Region [C]	45.36%
Region [D]	55.46%
Region [E]	68.24%
Region [F]	43.83%
Region [G]	21.53%
Region [H]	31.84%
Region [I]	70.29%
Region [J]	23.23%
Region [K]	41.71%
Region [L]	7.84%

There are graphs according to the forecasting results for plastic waste obtained

from ARIMA in Figures 41, 42 and 43. These graphs belong to the 3 regions where the best results were obtained, the other regions' graphics are in Appendix H of the thesis. When examined by region, the 3 regions with the best results are "Region L", "Region J", and "Region B ", respectively. MAPE values were 8.05%, 22.66% and 29.14% respectively. The model was run by selecting parameters (0,0,0) for Region L, parameters (0,0,1) for Region J and parameters (1,0,1) for Region B.

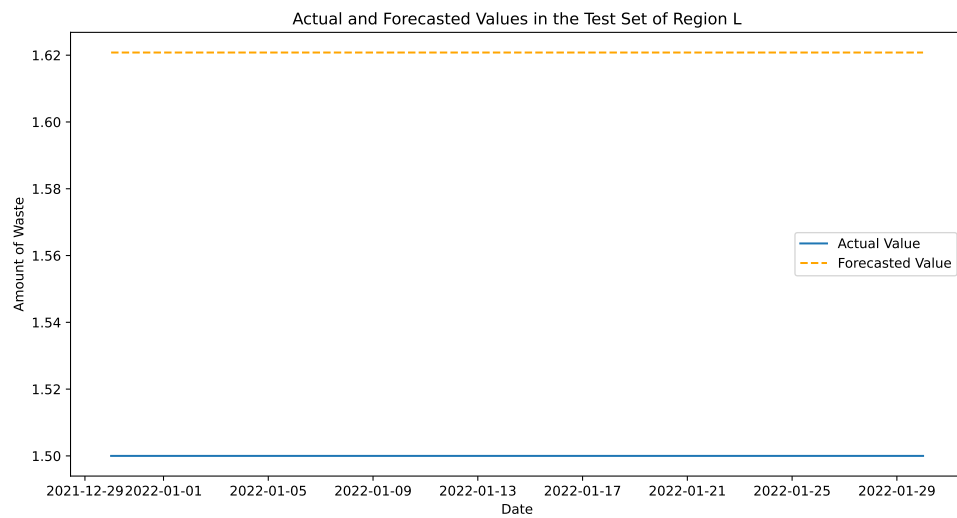


Figure 41. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region L

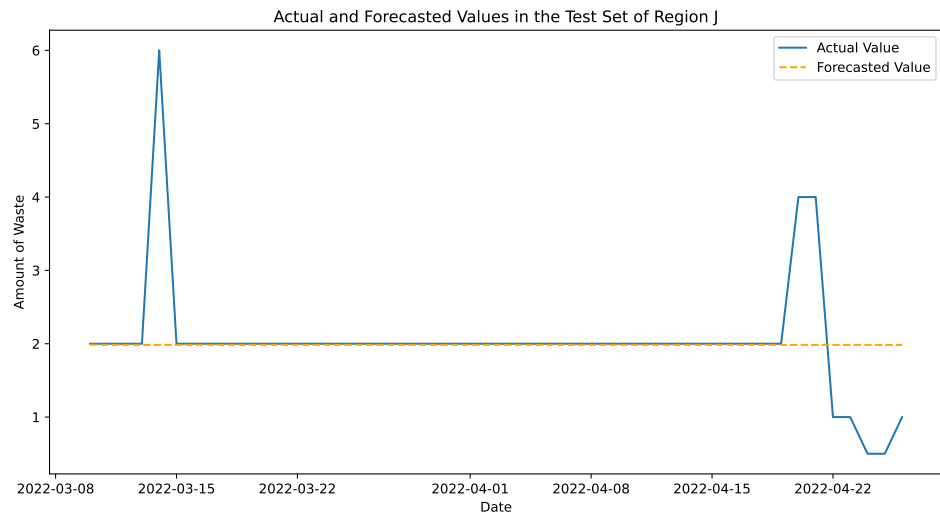


Figure 42. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region J

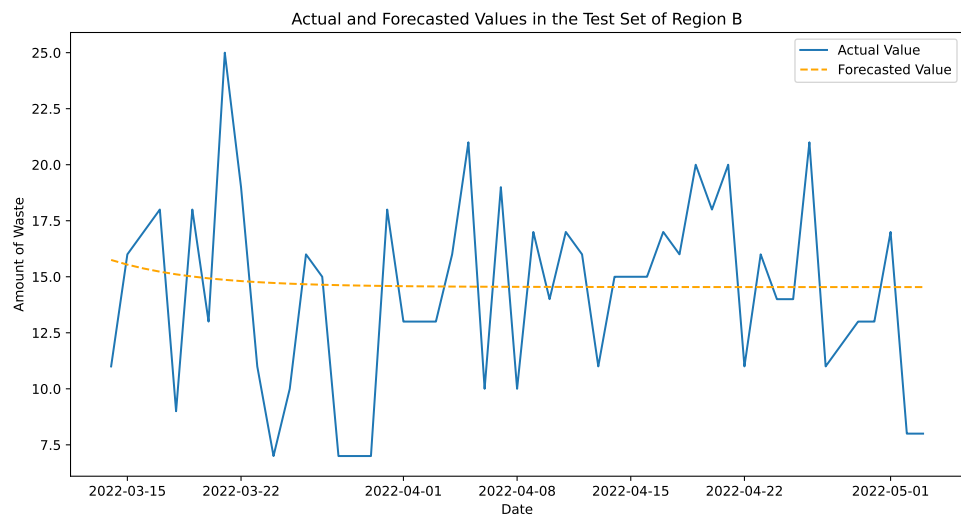


Figure 43. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region B

Table 29. Plastic Waste Forecasting Results Obtained from Testset using ARIMA

Region Name	MSE	MAPE	RMSE
Region [A]	1.06	39.74%	1.03
Region [B]	17.23	29.14%	4.15
Region [C]	3.36	64.47%	1.83
Region [D]	16.44	50.34%	4.05
Region [E]	98.10	276.78%	9.90
Region [F]	2.80	59.32%	1.67
Region [G]	0.45	50.90%	0.67
Region [H]	2.09	70.23%	1.45
Region [I]	1.02	57.97%	1.01
Region [J]	0.66	22.66%	0.81
Region [K]	3.80	199.66%	1.95
Region [L]	0.01	8.05%	0.12

6.3.4. *Interpreting the Results of Supervised Machine Learning Models*

Forecast results between the applied machine learning models were evaluated using performance metrics. Depending on the structure of the machine learning models used, K-fold cross-validation and Hold-out methods were preferred to obtain the results. This thesis presented the average of 5 different results obtained using the K-fold cross-validation method in XGBoost and KNN models. On the other hand, the Hold-out method was used in the ARIMA model and the results were found.

For paper waste data, it is observed that the ARIMA model is more successful than the other two models (K-NN and XGBoost) in terms of MSE value in 7 regions. While the KNN model was more successful in 2 regions, a low MSE value was obtained in the XGBoost model in 3 regions. Table 30 contains MSE, RMSE and MAPE values for paper waste. When comparing between regions, the lowest MAPE results in the XGBoost model were obtained from regions A, C and H, the K-NN model was obtained from regions D, H and I, and the ARIMA model was obtained from regions I, J and L.

Table 30. Comparison of Consistencies of Machine Learning Models for Paper Waste

	XGBOOST			KNN			ARIMA		
	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE
REGION A	666	51.24%	25.80	577	48.53%	24.02	725	73.26%	26.9
REGION B	1496	66.25%	38.68	1169	62.55%	34.19	1139	50.33%	33.7
REGION C	1264	48.10%	35.56	1394	66.30%	37.34	1369	70.61%	37
REGION D	5123	50.56%	71.58	4414	48.27%	66.44	3946	52.84%	62.8
REGION E	5572	135.57%	74.65	3879	119.40%	62.28	2395	66.41%	48.9
REGION F	2796	109.25%	52.88	2013	89.89%	44.87	2341	66.87%	48.3
REGION G	622	67.06%	24.93	672	74.65%	25.92	773	68.15%	27.8
REGION H	450	47.41%	21.22	395	49.70%	19.88	417	46.89%	20.4
REGION I	247	53.10%	15.73	236	51.14%	15.37	163	31.39%	12.8
REGION J	243	144.31%	15.59	182	135.66%	13.50	105	42.69%	10.2
REGION K	1586	73.46%	39.83	999	78.67%	31.60	669	63.9%	25.9
REGION L	22015	148.05%	148.37	17824	141.33%	133.51	5509	29.99%	74.2

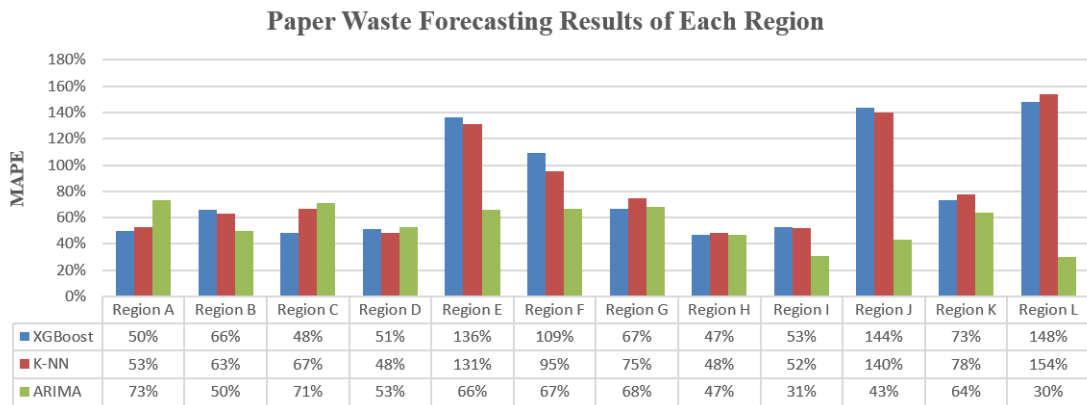


Figure 44. Paper Waste Forecasting Results of Each Region

It is observed that the ARIMA model for plastic waste data is more successful than the other two models in terms of MSE value in 9 regions. In 3 regions, a more successful MSE value was obtained than the KNN model. Table 31 contains MSE, RMSE and MAPE values for plastic waste. When compared between regions, the lowest MAPE results were obtained from G, J and I regions in the XGBoost and KNN models, and from G, J and L regions in the ARIMA model. The ARIMA model with the lowest MSE value for both waste types measures how close the forecasts are to the

actual values.

Table 31. Comparison of Consistencies of Machine Learning Models for Plastic Waste

	XGBOOST			K-NN			ARIMA		
	MSE	MAPE	RMSE	MSE	MAPE	RMSE	MSE	MAPE	RMSE
REGION A	4.85	51.36%	2.20	3.69	46.55%	1.92	1.06	39.74%	1.03
REGION B	49.32	46.60%	7.02	40.24	42.62%	6.34	17.23	29.14%	4.15
REGION C	2.99	48.11%	1.73	2.88	58.57%	1.70	3.36	64.47%	1.83
REGION D	19.65	60.95%	4.43	15.61	70.25%	3.95	16.44	50.34%	4.05
REGION E	121.42	263.32%	11.02	48.37	179.19%	6.95	98.10	276.78%	9.90
REGION F	3.56	73.67%	1.89	3.63	84.22%	1.91	2.80	59.32%	1.67
REGION G	0.83	36.82%	0.91	0.70	40.39%	0.84	0.45	50.90%	0.67
REGION H	7.24	132.41%	2.69	6.04	127.92%	2.46	2.09	70.23%	1.45
REGION I	1.58	118.21%	1.26	1.22	104.30%	1.11	1.02	57.97%	1.01
REGION J	1.21	42.69%	1.10	0.73	33.32%	0.85	0.66	22.66%	0.81
REGION K	10.94	108.21%	3.31	7.11	93.99%	2.67	3.80	199.66%	1.95
REGION L	34.05	163.89%	5.84	1.39	14.32%	1.18	0.01	8.05%	0.12

6.4. Comparison of Machine Learning and Traditional Methods

In this study, MSE value, which is one of the performance metrics, is used to find the method that gives the best results among the methods. Among traditional forecasting methods, MA, and machine learning models, ARIMA gives the best results. In this section, the Mann-Whitney U test is used to determine whether there is a significant difference between the results obtained from two different methods. This non-parametric test is particularly useful when the assumptions of parametric tests, such as the *t*-test, are not met or when the data is not normally distributed (McKnight and Najab, 2010). In the context of comparing two forecasting methods, such as ARIMA and MA, the Mann-Whitney U test evaluates whether there is a statistically significant difference in the performance of these methods in terms of forecasting accuracy. The test works by comparing the ranks of observations from each group. It calculates a test statistic based on the sum of ranks of observations from one group relative to the other. The resulting test statistic is then compared to a critical value or converted into a *p*-value, which represents the probability of observing such an extreme

result under the null hypothesis (i.e., no difference between the two methods). If the p -value is less than a predetermined significance level (commonly set at 0.05), it suggests that there is sufficient evidence to reject the null hypothesis and conclude that there is a statistically significant difference between the two methods. Conversely, if the p -value is greater than the significance level, it indicates that there is not enough evidence to reject the null hypothesis, and no significant difference between the methods can be inferred. Therefore, the Mann-Whitney U test provides a robust statistical method for comparing the performance of different forecasting methods and determining whether one method outperforms the other in terms of forecast accuracy (McKnight and Najab, 2010). To reach a general conclusion, the following hypothesis is established to determine whether there is a statistically significant difference between the MA and ARIMA models.

- Null hypothesis (H_0): There is no significant difference between the two methods.
- Alternative hypothesis (H_1): There is a significant difference between the two methods.

For the Mann-Whitney U test, the `mannwhitneyu` function in the `scipy.stats` module is used in the Python code. The "mannwhitneyu" function is imported from the "scipy.stats" module. This function performs the Mann-Whitney U test and calculates the test statistic and p -value. As a significance level α , 0.05, is determined. The Mann-Whitney U test contains a test statistic that expresses the difference between the two groups. This statistic represents the difference in ordered totals between groups. 12 MSE values are obtained both in two methods. When this method was applied between the results of the two methods, the test statistic value was 90. The p -value refers to the result of hypothesis testing in the Mann-Whitney U test. This value measures the accuracy of the null hypothesis. The null hypothesis generally states that there is no statistically significant difference between the two groups. In this case, the p -value was calculated as 0.31. Compared to the established significance level ($\alpha = 0.05$), the p -value is 0.31. Since this value is greater than the determined significance level (0.05), the null hypothesis (H_0) cannot be rejected. That is, there is no statistically significant

difference between the two groups. As a result, according to the Mann-Whitney U test results, it was concluded that there was no significant difference between the two different methods in the data sets. This indicates that there is no statistically significant difference between the ARIMA and MA methods. In other words, the results of both methods are successful and can be used to reach the best prediction result.

In Table 32 among all methods for paper waste, the ARIMA model gives the lowest result in region L. For the other regions, the MA gives the lowest MSE value.

Table 32. Comparison of All Methods for Paper Waste

PAPER WASTE MSE VALUES							
	MA	WMA	ES	HW	XGBOOST	K-NN	ARIMA
REGION A	357	385	590	517	666	577	725
REGION B	615	715	1126	1110	1496	1169	1139
REGION C	765	856	1247	1008	1264	1394	1369
REGION D	2532	2764	4168	4103	5123	4414	3946
REGION E	1527	1599	2684	3167	5572	3879	2395
REGION F	810	859	1422	1621	2796	2013	2341
REGION G	317	333	535	538	622	672	773
REGION H	172	194	309	288	450	395	417
REGION I	108	114	183	204	247	236	163
REGION J	97	103	163	165	243	182	105
REGION K	472	564	972	964	1586	999	669
REGION L	5885	6637	11136	13237	22015	17824	5509

For plastic waste, the MA gives the lowest MSE value in 10 regions. In 2 regions, the ARIMA model gives the lowest result in Table 33.

Table 33. Comparison of All Methods for Plastic Waste

PLASTIC WASTE MSE VALUES							
	MA	WMA	ES	HW	XGBOOST	K-NN	ARIMA
REGION A	1.9	2.39	4.4	3.6	4.85	3.69	1.06
REGION B	16	17.08	28.5	30.1	49.32	40.24	17.23
REGION C	1.6	1.73	2.6	2.0	2.99	2.88	3.36
REGION D	7.7	8.34	13.4	31	19.65	15.61	16.44
REGION E	16.1	17.88	29.6	41	121.42	48.37	98.10
REGION F	0.99	1.2	2.02	2.6	3.56	3.63	2.80
REGION G	0.25	0.29	0.49	0.56	0.83	0.70	0.45
REGION H	0.78	1.1	1.48	3.9	7.24	6.04	2.09
REGION I	0.4	0.54	0.87	1.1	1.58	1.22	1.02
REGION J	0.39	0.4	0.68	0.68	1.21	0.73	0.66
REGION K	2.02	2.91	4.6	15.3	10.94	7.11	3.80
REGION L	0.13	0.67	1.16	1.2	34.05	1.39	0.01

CHAPTER 7: AUTOMATION FRAMEWORK FOR FORECASTING WASTES

This chapter aims to create a code structure that will enable the results obtained from the dataset and forecasting models to be obtained simultaneously. Firstly, traditional forecasting methods are run and those that give the best results according to the determined parameters are shown. Secondly, the best forecast results according to machine learning models are presented according to the parameters and written in the same file. Based on all the results obtained, an algorithm is written that finds the best result among all models. Details about the algorithm are included in Section 7.1.

7.1. *Analysis Steps*

The analysis of the automation framework for forecasting waste amounts involves several key steps, as depicted in Figure 45. Initially, the dataset is loaded and prepared for analysis in Step 1. Following this in Step 2, both traditional and machine learning approaches for forecasting are processed in parallel in the training phase. Within the traditional forecasting methods (Step 2a), four models are developed: Moving Average (MA) (Step 2a-1), Weighted Moving Average (WMA) (Step 2a-2), Exponential Smoothing (ES) (Step 2a-3), and Holt-Winters (HW) (Step 2a-4). Similarly, within the machine learning approaches (Step 2b), models are established for K-Nearest Neighbors (K-NN) (Step 2b-1), Extreme Gradient Boosting (XGBoost) (Step 2b-2) and Autoregressive Integrated Moving Average (ARIMA (Step 2b-3). Subsequently, parameters are determined for each of the models. In Step 3, various methods can be applied according to the literature during the parameter selection process. For instance, optimal value selection for the parameter selection process of MA can be demonstrated as an example by incrementally increasing p period according to a certain increment rate (Jasemi and Kimiagari, 2012). In WMA, this process involves the addition of expert domain knowledge, where specific weights are assigned to specific days Shih and Tsokos (2008). In ES, the α value can be determined through grid search, where different α values are tried, and the α that provides

the best performance is selected (Tratar et al., 2016). In Holt-Winters, grid search is utilized to determine the optimal values for α , β , and γ parameters by systematically testing various combinations to find the best fit for the data (Jiang et al., 2020). For parameter selection in K-NN models, techniques like grid search and randomized search are employed to optimize hyperparameters such as $n_neighbors$, $distance\ metrics$, and $weights$ (Belete and Huchaiah, 2022). For parameter selection in forecasting with XGBoost, techniques like grid search and randomized search are applied to optimize hyperparameters such as $n_estimators$, max_depth , and $learning\ rate$, as well as regularization parameters like α and λ (Yang et al., 2021), (Javid et al., 2023). In ARIMA, parameter selection can be facilitated using methods such as Auto ARIMA with AIC (*Akaike Information Criterion*) and BIC (*Bayesian Information Criterion*), focusing on determining the appropriate values for the p , d , and q hyperparameters (Choudhary et al., 2022), (Meisenbacher et al., 2022). Step 4 is to analyze the best model according to the parameters that provide the best results for each approach. Finally, Step 5 reports the best model with the most accurate forecasting performance among other models.

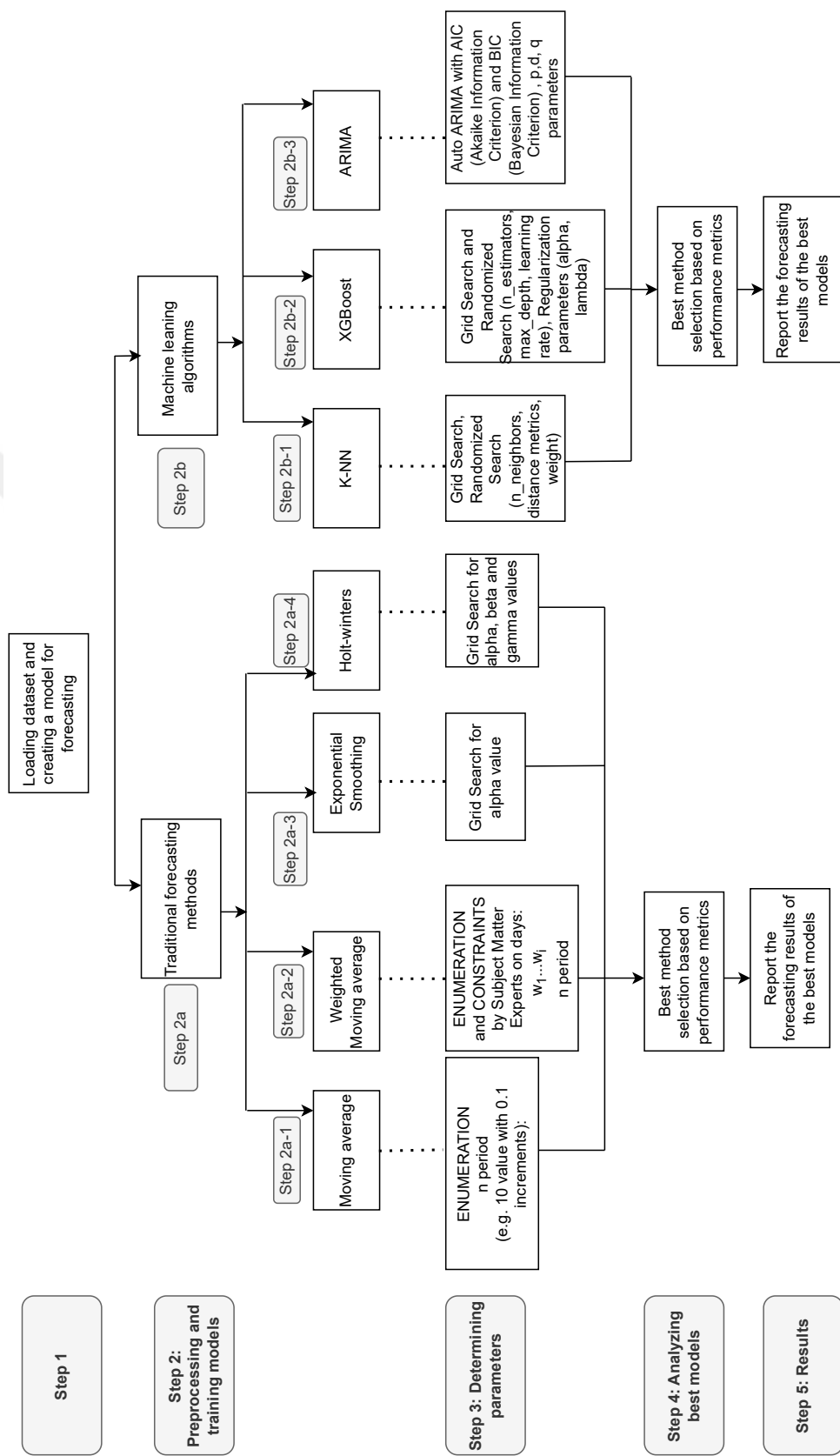


Figure 45. Steps of Analysis of Automation Framework for Forecasting

Algorithm 1 Automation Framework for Forecasting

- 1: Prepare the data:
 - 2: Divide the data into train and test sets
 - 3: $train$ = first 80% of data
 - 4: $test$ = remaining 20% of data
 - 5: Create Moving Average model:
 - 6: ma_window = Choose appropriate window size for Moving Average
 - 7: $ma_forecast$ = Calculate Moving Average forecast for the entire dataset
 - 8: Create Weighted Moving Average model:
 - 9: wma_window = Choose appropriate window size for Weighted Moving Average
 - 10: $ma_forecast$ = Calculate Weighted Moving Average forecast for the entire dataset
 - 11: Create Exponential Smoothing model:
 - 12: es_alpha = Choose the appropriate smoothing factor ($alpha$) for Exponential Smoothing based on Grid Search
 - 13: es_fit = Fit Exponential Smoothing model to the entire dataset
 - 14: Create Holt-Winters model:
 - 15: es_alpha = Choose the appropriate parameters for Holt-Winters based on Grid Search
 - 16: hw_fit = Fit Holt-Winters model to the entire dataset
 - 17: Create XGBoost model:
 - 18: xgb_model = Initialize XGBRegressor with the best parameters obtained from Grid Search
 - 19: kf = Create a 5-fold cross-validation
 - 20: xgb_mse_scores = Evaluate xgb_model with cross-validation using kf
 - 21: xgb_avg_mse = Average the MSE scores
-

Algorithm 1 Automation Framework for Forecasting (continued)

- 1: Create KNeighborsRegressor model:
 - 2: knn_model = Initialize KNeighborsRegressor with the best K value obtained from Grid Search
 - 3: kf = Create a 5-fold cross-validation
 - 4: knn_mse_scores = Evaluate knn_model with cross-validation using kf
 - 5: knn_avg_mse = Average the MSE scores
 - 6: Create ARIMA model:
 - 7: $arima_model$ = Initialize ARIMA model with the best hyperparameters obtained from *auto-arima*
 - 8: $arima_fit$ = Fit $arima_model$ to *train* data
 - 9: $arima_forecast$ = Forecast $arima_fit$ for the length of *test*
 - 10: $arima_forecast_series$ = Convert forecast to a pandas Series with index as *test.index*
 - 11: $arima_forecast_test$ = Select forecast values for *test*
 - 12: $arima_avg_mse$ = Average the MSE scores
 - 13: Determine the best model based on MSE:
 - 14: $mse_results = \{ 'Moving\ Average': Calculate\ MSE \}$
 - 15: $mse_results = \{ 'Weighted\ Moving\ Average': Calculate\ MSE \}$
 - 16: $mse_results = \{ 'Exponential\ Smoothing': Calculate\ MSE \}$
 - 17: $mse_results = \{ 'Holt-Winters': Calculate\ MSE \}$
 - 18: $mse_results = \{ 'XGBoost': xgb_avg_mse \}$
 - 19: $mse_results = \{ 'K-NN': knn_avg_mse, 'ARIMA': arima_mse \}$
 - 20: $mse_results = \{ 'ARIMA': arima_mse \}$
 - 21: $best_model$ = Identify model with the minimum MSE in $mse_results$
-

Algorithm 1 Automation Framework for Forecasting (continued)

```
1: Print the best model and its MSE:
2:   Print "Best Model:", best_model
3:   Print "MSE value:", mse_results[best_model]
4: Plot the predictions for the best model:
5: if best_model is 'Moving Average' then
6:   Plot actual values and forecasted values of the Moving Average model on the
     entire dataset with labels and title
7: else if best_model is 'Weighted Moving Average' then
8:   Plot actual values and forecasted values of the Weighted Moving Average
     model on the entire dataset with labels and title
9: else if best_model is 'Exponential Smoothing' then
10:  Plot actual values and forecasted values of the Exponential Smoothing model
     on the entire dataset with labels and title
11: else
12:  Plot actual values and forecasted values of Holt-Winters model on the entire
     dataset with labels and title
13: end if
```

Algorithm 1 Automation Framework for Forecasting (continued)

```
1: Print the best model and its MSE:
2:   Print "Best Model:", best_model
3:   Print "MSE value:", mse_results[best_model]
4: Plot the forecasting for the best model:
5: if best_model is 'XGBoost' then
6:   Fit xgb_model to train data
7:   y_pred = Forecast xgb_model on test data
8:   Scatter plot of test vs y_pred with labels and title
9: else if best_model is 'K-NN' then
10:  Fit knn_model to train data
11:  y_pred = Forecast knn_model on test data
12:  Scatter plot of test vs y_pred with labels and title
13: else
14:   Plot actual values and forecasted values of ARIMA model on test.index with
      labels and title
15: end if
```

CHAPTER 8: CONCLUSION

This study presents many different forecast methods that are applied to 2 types of waste datasets for 12 regions in Izmir, Turkey. The two types of recyclable wastes considered are paper and plastic recyclable wastes and they cover a significant portion of overall collected recyclable wastes. The main purpose of this thesis is to examine waste forecast processes that play an important role in the field of waste management and to contribute to various decision-making processes. Waste forecasting can be used in inventory management and production planning processes by forecasting waste generation and a facility's waste stock. These forecasts can provide valuable insights to a variety of organizations, such as revenue opportunities for recycling initiatives and facilitating specific budget planning. The research employs a comprehensive approach, starting with traditional forecasting methods. Subsequently, supervised machine learning techniques are employed to further enhance forecasting accuracy. In this thesis, supervised machine learning techniques such as XGBoost, K-NN and ARIMA are used to forecast waste amounts. While XGBoost is preferred for fast and accurate forecasts, K-NN provides simplicity and flexibility. ARIMA, on the other hand, is effective for time series data. All of these methods are examined and compared in forecasting waste amounts.

The forecast results obtained according to the traditional forecasting methods used were evaluated and the best and worst results of each method were presented by region. According to the results of the MA method, region I gave better results in MAPE value with 27% and region J gave the worst results with 76% for paper waste. Region L gave a better results value with 2%, and region E gave the worst results with 43% for plastic waste. WMA method resulted as follows, region I gave better results in MAPE value with 40% and region J gave the worst results with 141% for paper waste. Region L gave a better results value with 2%, and region I gave the worst results with 90% for plastic waste. ES method resulted as follows, region I gave better results in MAPE value with 38% and region J gave the worst results with 138% for paper waste. Region L gave a better results value with 6%, and region E gave the worst results with 76% for

plastic waste. HW method resulted as follows, region H gave better results in MAPE value with 39% and region J gave the worst results with 139% for paper waste. Region L gave a better results value with 3%, and region E gave the worst results with 138% for plastic waste. Among traditional methods, the MA method was found to give the best results with the lowest MSE values. Traditional forecasting methods reveal that the moving average technique yields the most accurate forecasts across all regions, with an MSE value of 97 for paper waste and 0.13 for plastic waste, outperforming other methods.

According to the forecasting results obtained from the machine learning approaches used, the best and worst results of each method are presented by region. XGBoost algorithm resulted as follows, for paper waste region H gave the best result in MAPE value with 47% and the worst result was obtained from region L. For plastic waste region G gave the best result in MAPE value with 37% and the worst result was obtained from region E. The K-NN algorithm resulted as follows, for paper waste region D gave the best result in MAPE value with 48% and the worst result was obtained from region L. For plastic waste region L gave the best result in MAPE value with 14% and the worst result was obtained from region E. ARIMA algorithm resulted as follows, for paper waste region D gave the best result in MAPE value with 48% and the worst result was obtained from region A. For plastic waste region L gave the best result in MAPE value with 8% and the worst result was obtained from region E. Among machine learning methods, ARIMA gave the best results with the lowest MSE values. Among supervised machine learning methods indicate that the ARIMA approach generally delivers superior forecasting outcomes across the majority of regions, achieving MSE value of 105 for paper waste. These results are obtained with the current dataset; better results can be achieved with a larger dataset

The effectiveness of the methods used in analyzing and forecasting the amount of recyclable waste was examined. It was conducted as a case study based on the analysis of data sets containing the amounts of paper and plastic waste collected from different points in the Izmir Region. This study focuses on how waste forecast results can contribute to the waste collection process. In particular, it investigates the potential of forecast results of waste datasets obtained from different regions to optimize recycling

processes. In this context, the thesis includes the comparison of traditional forecasting methods and machine learning algorithms and determines which methods are more effective in forecasting waste amounts. In conclusion, this thesis makes a scientific contribution to the development of waste management strategies and sustainability efforts.

In terms of future studies, several important research areas can be identified based on the results of this thesis. First of all, more comprehensive research can be conducted in the field of waste management, and it can contribute to the development of a general waste management strategy, especially by focusing on the analysis and forecasting of waste amounts in different cities. Additionally, a more detailed comparison of different machine learning algorithms and traditional forecasting methods can be performed, allowing a more accurate forecasting of waste quantities. However, focus can be placed on the development of new and innovative forecasting methods to optimize waste management processes, allowing waste quantities to be managed more effectively. Finally, further work can be done on automated forecasting frameworks, which can combine different forecasting algorithms to choose the method that gives the best forecasting result and enables more precise forecasting of waste quantities. In this way, future research in the field of waste management can contribute to a better understanding and more effective management of waste quantities.

REFERENCES

- Abdella Ahmed, A. K., Ibraheem, A. M. and Abd-Ellah, M. K. (2022). *Forecasting of municipal solid waste multi-classification by using time-series deep learning depending on the living standard*, Results in Engineering, Vol. 16, pp. 100655. <https://www.sciencedirect.com/science/article/pii/S2590123022003255>
- Afrinaldi, F. (2020). *Exploring product lifecycle using markov chain*, Procedia Manufacturing, Vol. 43, pp. 391 – 398. Sustainable Manufacturing - Hand in Hand to Sustainability on Globe: Proceedings of the 17th Global Conference on Sustainable Manufacturing. <http://www.sciencedirect.com/science/article/pii/S2351978920307824>
- Barrow, D. and Kourentzes, N. (2018). *The impact of special days in call arrivals forecasting: A neural network approach to modelling special days*, European Journal of Operational Research, Vol. 264 (3), pp. 967–977. <https://www.sciencedirect.com/science/article/pii/S0377221716305525>
- Belete, D. M. and Huchaiah, M. D. (2022). *Grid search in hyperparameter optimization of machine learning models for prediction of hiv/aids test results*, International Journal of Computers and Applications, Vol. 44 (9), pp. 875–886. <https://doi.org/10.1080/1206212X.2021.1974663>
- Birgen, C., Magnanelli, E., Carlsson, P., Øyvind Skreiberg, Mosby, J. and Becidan, M. (2021). *Machine learning based modelling for lower heating value prediction of municipal solid waste*, Fuel, Vol. 283, pp. 118906. <https://www.sciencedirect.com/science/article/pii/S0016236120319025>
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system, pp. 785–794.
- Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E. O., Waser, L. T., Travaglini, D. and McRoberts, R. E. (2016). *A meta-analysis and review of the literature on the k-nearest neighbors technique for forestry applications that use remotely sensed data*, Remote Sensing of Environment, Vol. 176, pp. 282–294. <https://www.sciencedirect.com/science/article/pii/S0034425716300293>
- Choudhary, A., Kumar, S., Sharma, M. and Sharma, K. P. (2022). *A framework*

for data prediction and forecasting in wsn with auto arima, Wireless Personal Communications, Vol. pp. 1–15.

Das, K., Kumar, R. and Krishna, A. (2024). *Analyzing electric vehicle battery health performance using supervised machine learning*, Renewable and Sustainable Energy Reviews, Vol. 189, pp. 113967. <https://www.sciencedirect.com/science/article/pii/S1364032123008250>

Garre, A., Ruiz, M. C. and Hontoria, E. (2020). *Application of machine learning to support production planning of a food industry in the context of waste generation under uncertainty*, Operations Research Perspectives, Vol. 7, pp. 100147. <https://www.sciencedirect.com/science/article/pii/S2214716019301988>

Hong, Z., Tao, M., Liu, L., Zhao, M. and Wu, C. (2023). *An intelligent approach for predicting overbreak in underground blasting operation based on an optimized xgboost model*, Engineering Applications of Artificial Intelligence, Vol. 126, pp. 107097. <https://www.sciencedirect.com/science/article/pii/S0952197623012812>

Hoque, M. M. and Rahman, M. T. U. (2020). *Landfill area estimation based on solid waste collection prediction using ann model and final waste disposal options*, Journal of Cleaner Production, Vol. 256, pp. 120387. <https://www.sciencedirect.com/science/article/pii/S0959652620304340>

Hossain, M., Santhanam, A., Norulaini, N. and Kadir, M. (2011). *Clinical solid waste management practices and its impact on human health and environment - a review*, Waste management (New York, N.Y.), Vol. 31, pp. 754–66.

Hyndman, R. J. and Athanasopoulos, G. (1998). *Forecasting: Principles and Practice*, 3rd Edition, Monash University.

Jasemi, M. and Kimiagari, A. M. (2012). *An investigation of model selection criteria for technical analysis of moving average*, Journal of Industrial Engineering International, Vol. 8, pp. 1–9.

Javid, I., Ghazali, R., Batool, T., Jafri, S. I. H. and Altaf, A. (2023). *An arima and xgboost model utilized for forecasting municipal solid waste generation*, *International Conference on Information, Communication and Computing Technology*, Springer, pp. 16–28.

Jiang, W., Wu, X., Gong, Y., Yu, W. and Zhong, X. (2020). *Holt–winters*

smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption, Energy, Vol. 193, pp. 116779.

Klein, L., Zelinka, I. and Seidl, D. (2024). *Optimizing parameters in swarm intelligence using reinforcement learning: An application of proximal policy optimization to the isoma algorithm*, Swarm and Evolutionary Computation, Vol. 85, pp. 101487. <https://www.sciencedirect.com/science/article/pii/S2210650224000208>

Law, A. M. and Kelton, W. D. (2007). *Simulation Modeling and Analysis*, 4th Edition, McGraw-Hill.

Marcot, B. G. and Hanea, A. M. (2021). *What is an optimal value of k in k-fold cross-validation in discrete bayesian network analysis?*, Computational Statistics, Vol. 36, pp. 1613–9658.

Mazumder, S., Neogy, S. and Das, S. (2023). *Simple moving average (sma) crossover strategy with buy sell indicator*, Asia-Pacific Journal of Management and Technology, Vol. 03, pp. 26–40.

McKnight, P. E. and Najab, J. (2010). *Mann-whitney u test*, The Corsini encyclopedia of psychology, Vol. pp. 1–1.

Meisenbacher, S., Turowski, M., Phipps, K., Rätz, M., Müller, D., Hagenmeyer, V. and Mikut, R. (2022). *Review of automated time series forecasting pipelines*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, Vol. 12 (6), pp. e1475.

Mitchell, T. M. (1997). *Machine Learning*, McGraw Hill.

Muneeb, S. M., Adhami, A. Y., Jalil, S. A. and Asim, Z. (2018). *Decentralized bi-level decision planning model for municipal solid waste recycling and management with cost reliability under uncertain environment*, Sustainable Production and Consumption, Vol. 16, pp. 33–44. <https://www.sciencedirect.com/science/article/pii/S2352550918301660>

Oguz-Ekim, P. (2021). *Machine learning approaches for municipal solid waste generation forecasting*, Environmental Engineering Science, Vol. 38 (6), pp. 489–499. <https://doi.org/10.1089/ees.2020.0232>

Ostertagova, E. and Ostertag, O. (2012). *Forecasting using simple exponential smoothing method*, Acta Electrotechnica et Informatica, Vol. 12 (3), pp. 62.

Paulauskaite-Taraseviciene, A., Raudonis, V. and Sutiene, K. (2022).

- Forecasting municipal solid waste in lithuania by incorporating socioeconomic and geographical factors*, Waste Management, Vol. 140, pp. 31–39.
<https://www.sciencedirect.com/science/article/pii/S0956053X22000046>
- Perry, M. (2010). *The Weighted Moving Average Technique*.
- Pillai, V. M. and Chandrasekharan, M. (2008). *An absorbing markov chain model for production systems with rework and scrapping*, Computers & Industrial Engineering, Vol. 55 (3), pp. 695 – 706.
<http://www.sciencedirect.com/science/article/pii/S036083520800051X>
- Punyapornwithaya, V., Arjkumpa, O., Buamithup, N., Kuatako, N., Klaharn, K., Sansamur, C. and Jampachaisri, K. (2023). *Forecasting of daily new lumpy skin disease cases in thailand at different stages of the epidemic using fuzzy logic time series, nnar, and arima methods*, Preventive Veterinary Medicine, Vol. 217, pp. 105964.
<https://www.sciencedirect.com/science/article/pii/S0167587723001289>
- Qambar, A. S. and Al Khalidy, M. M. M. (2023). *Development of local and global wastewater biochemical oxygen demand real-time prediction models using supervised machine learning algorithms*, Engineering Applications of Artificial Intelligence, Vol. 118, pp. 105709.
<https://www.sciencedirect.com/science/article/pii/S0952197622006996>
- Razzaq, A., Sharif, A., Najmi, A., Tseng, M.-L. and Lim, M. K. (2021). *Dynamic and causality interrelationships from municipal solid waste recycling to economic growth, carbon emissions and energy efficiency using a novel bootstrapping autoregressive distributed lag*, Resources, Conservation and Recycling, Vol. 166, pp. 105372.
<https://www.sciencedirect.com/science/article/pii/S092134492030687X>
- Seçkiner, S., Shumye, A. M. and Geçer, S. (2021). *Minimizing solid waste collection routes using ant colony algorithm: A case study in gaziantep district*, Journal of Transportation and Logistics, Vol. 6 (1), pp. 29–47.
- Shih, S. H. and Tsokos, C. P. (2008). *A weighted moving average process for forecasting*, Journal of Modern Applied Statistical Methods, Vol. 7, pp. 187–197.
- Solak, S. G. and Pekküçükşen, S. (2018). *Türkiye’de kentsel katı atık yönetimi: Karşılaştırmalı bir analiz*, MANAS Sosyal Araştırmalar Dergisi, Vol. 7, pp. 0 – 0.
- Son, H., Ahn, J., Chung, A. D. and Drumwright, M. E. (2023). *From*

- the black box to the glass box: Using unsupervised and supervised learning processes to predict user engagement for the airline companies*, International Journal of Information Management Data Insights, Vol. 3 (2), pp. 100181. <https://www.sciencedirect.com/science/article/pii/S2667096823000289>
- Sunayana, Kumar, S. and Kumar, R. (2021). *Forecasting of municipal solid waste generation using non-linear autoregressive (nar) neural models*, Waste Management, Vol. 121, pp. 206–214. <https://www.sciencedirect.com/science/article/pii/S0956053X20307005>
- Tratar, L. F., Mojškerc, B. and Toman, A. (2016). *Demand forecasting with four-parameter exponential smoothing*, International Journal of Production Economics, Vol. 181, pp. 162–173.
- Tsiliyannis, C. A. (2018). *Markov chain modeling and forecasting of product returns in remanufacturing based on stock mean-age*, European Journal of Operational Research, Vol. 271 (2), pp. 474 – 489. <http://www.sciencedirect.com/science/article/pii/S0377221718304260>
- Twumasi, C. and Twumasi, J. (2022). *Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of tema general hospital of ghana*, International Journal of Forecasting, Vol. 38 (3), pp. 1258–1277. <https://www.sciencedirect.com/science/article/pii/S0169207021001710>
- Van der Walt, A. and Bean, W. L. (2022). *Inventory management for the in-flight catering industry: A case of uncertain demand and product substitutability*, Computers & Industrial Engineering, Vol. 165, pp. 107914. <https://www.sciencedirect.com/science/article/pii/S0360835221008184>
- Wan, A., Gong, Z., Chen, T. and AL-Bukhaiti, K. (2024). *Mass flow characteristics prediction of refrigerants through electronic expansion valve based on xgboost*, International Journal of Refrigeration, Vol. 158, pp. 345–352. <https://www.sciencedirect.com/science/article/pii/S0140700723004668>
- Wang, M., Hui, G., Pang, Y., Wang, S. and Chen, S. (2023). *Optimization of machine learning approaches for shale gas production forecast*, Geoenergy Science and Engineering, Vol. 226, pp. 211719. <https://www.sciencedirect.com/science/article/pii/S2949891023003068>

- Wu, D., Yuan, C., Kumfer, W. and Liu, H. (2017). *A life-cycle optimization model using semi-markov process for highway bridge maintenance*, Applied Mathematical Modelling, Vol. 43, pp. 45 – 60. <http://www.sciencedirect.com/science/article/pii/S0307904X16305534>
- Xu, H., Zhou, J., G. Asteris, P., Jahed Armaghani, D. and Tahir, M. M. (2019). *Supervised machine learning techniques to the prediction of tunnel boring machine penetration rate*, Applied Sciences, Vol. 9 (18). <https://www.mdpi.com/2076-3417/9/18/3715>
- Xu, Z., Elomri, A., Pokharel, S., Zhang, Q., Ming, X. and Liu, W. (2017). *Global reverse supply chain design for solid waste recycling under uncertainties and carbon emission constraint*, Waste Management, Vol. 64, pp. 358 – 370. <http://www.sciencedirect.com/science/article/pii/S0956053X17301022>
- Yang, L., Zhao, Y., Niu, X., Song, Z., Gao, Q. and Wu, J. (2021). *Municipal solid waste forecasting in china based on machine learning models*, Frontiers in Energy Research, Vol. 9, pp. 763977.
- Yin, R. K. (2013). *Case Study Research: Design and Methods*, Sage publications.
- Zhang, C., Dong, H., Geng, Y., Liang, H. and Liu, X. (2022). *Machine learning based prediction for china's municipal solid waste under the shared socio-economic pathways*, Journal of Environmental Management, Vol. 312, pp. 114918. <https://www.sciencedirect.com/science/article/pii/S0301479722004911>

APPENDICES

Appendix A - Histogram and Curve Fitting Line of Wastes in Regions

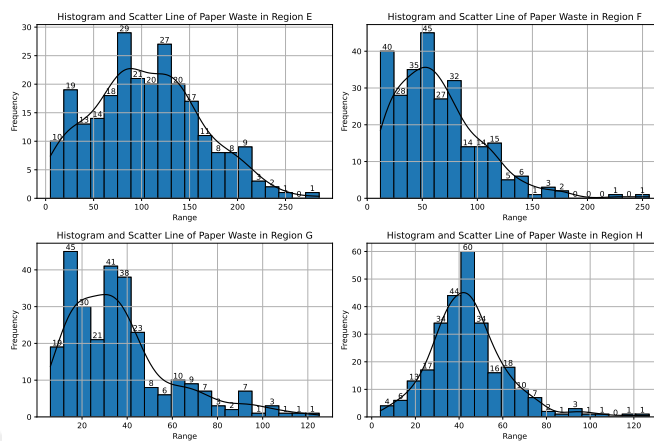


Figure 46. Histogram and Curve Fitting Line of Paper Waste in Regions E-H

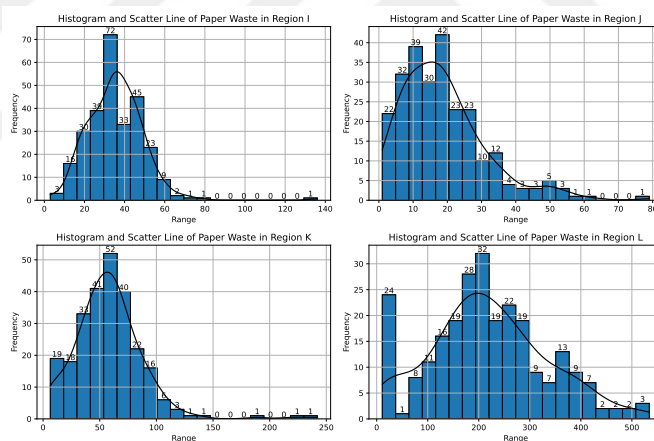


Figure 47. Histogram and Curve Fitting Line of Paper Waste in Regions I-L

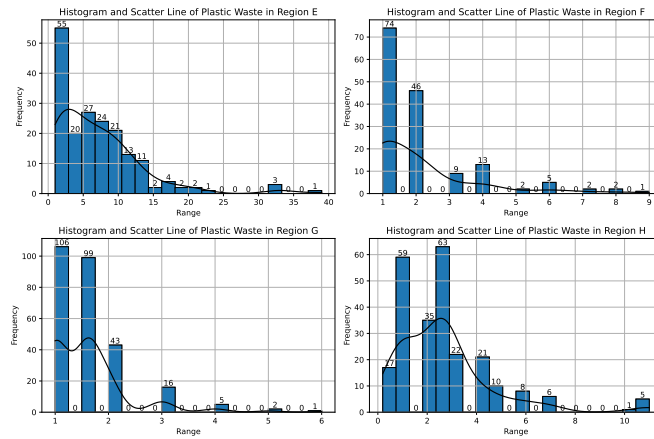


Figure 48. Histogram and Curve Fitting Line of Plastic Waste in Regions E-H

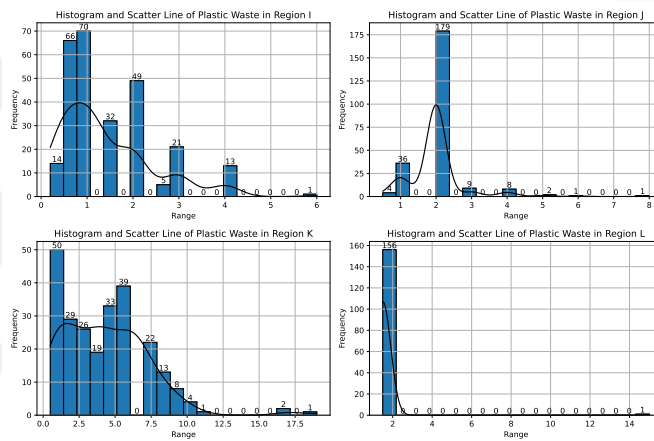


Figure 49. Histogram and Curve Fitting Line of Plastic Waste in Regions I-L

Appendix B - Moving Average Method for Wastes in Regions

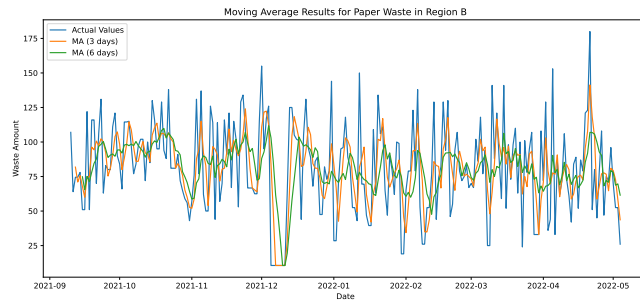


Figure 50. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region B

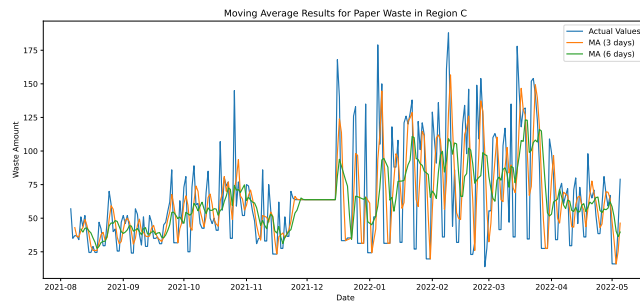


Figure 51. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region C

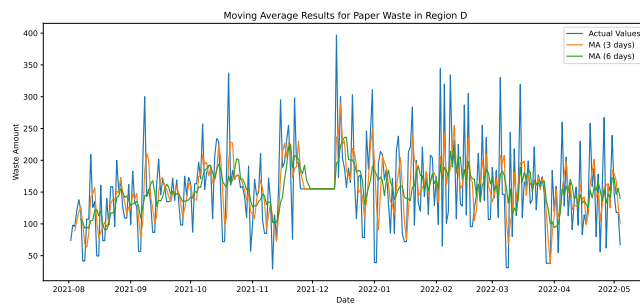


Figure 52. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region D

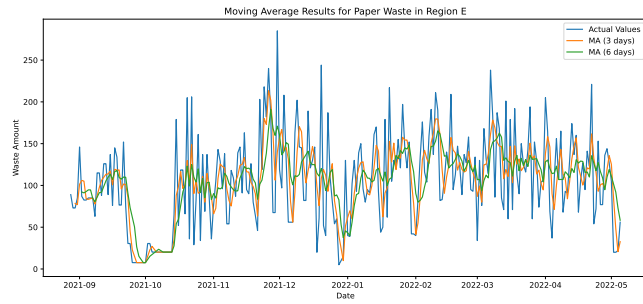


Figure 53. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region E

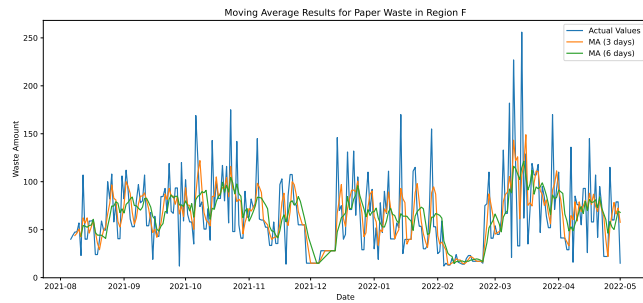


Figure 54. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region F

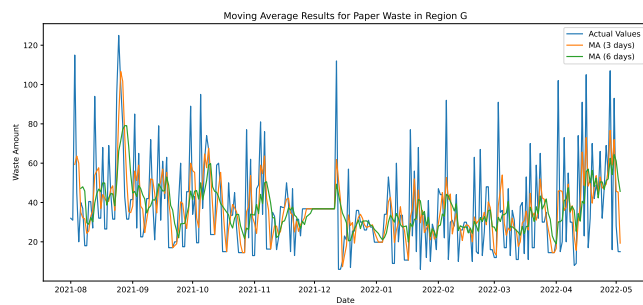


Figure 55. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region G

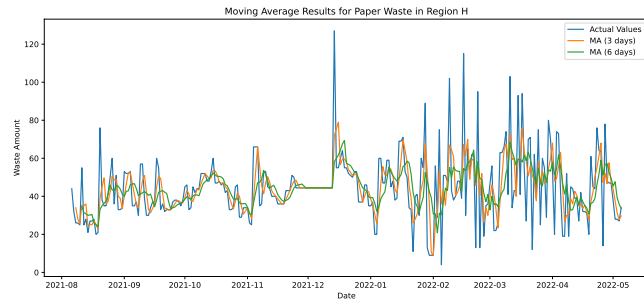


Figure 56. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region H

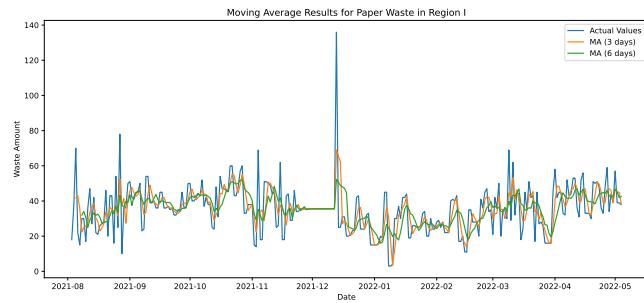


Figure 57. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region I

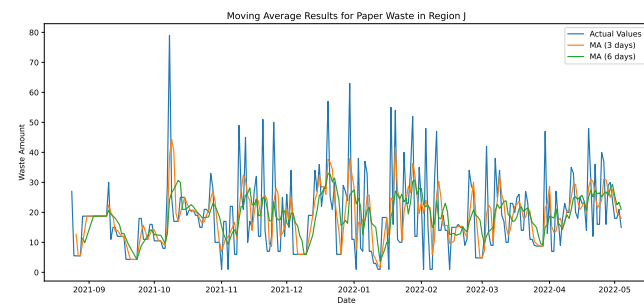


Figure 58. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region J

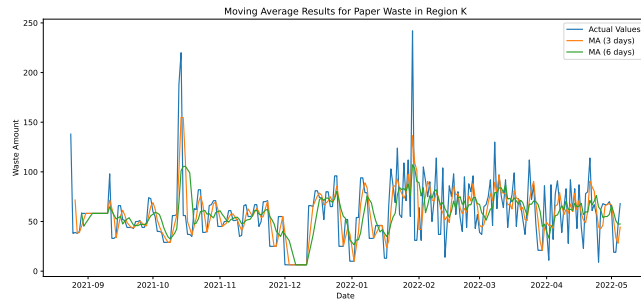


Figure 59. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region K

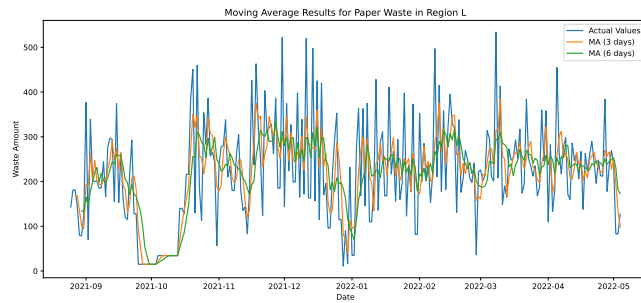


Figure 60. Moving Average Method Using 3 & 6 Days Period for Paper Waste in Region L

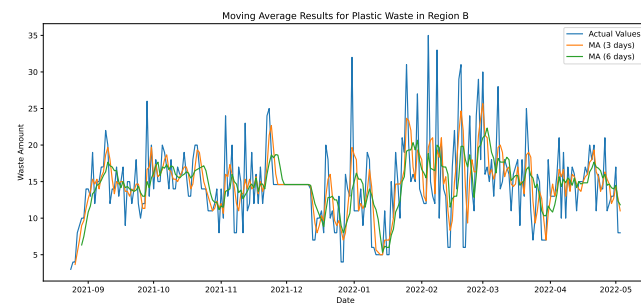


Figure 61. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region B

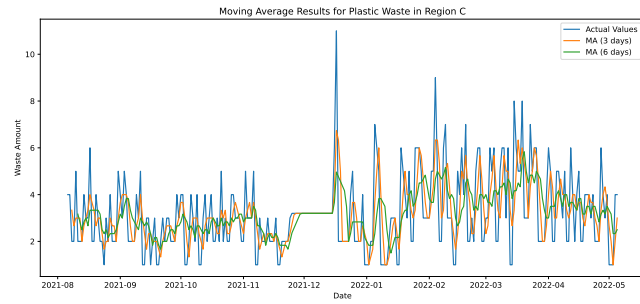


Figure 62. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region C

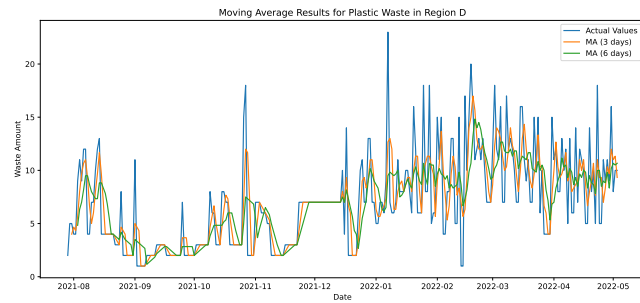


Figure 63. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region D

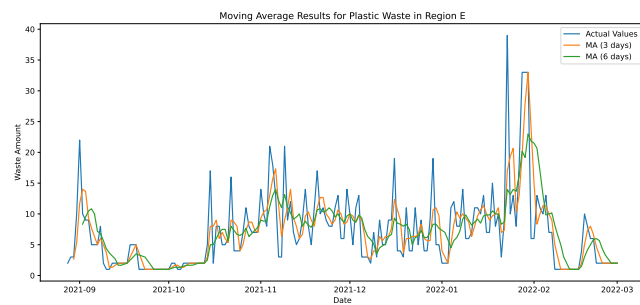


Figure 64. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region E

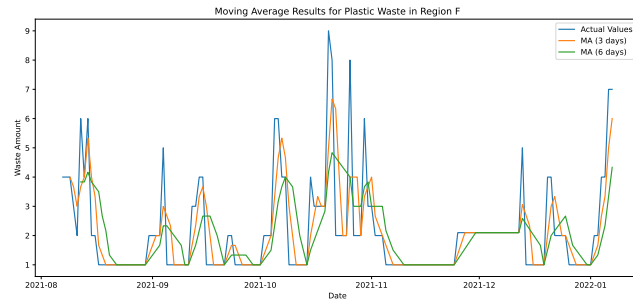


Figure 65. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region F

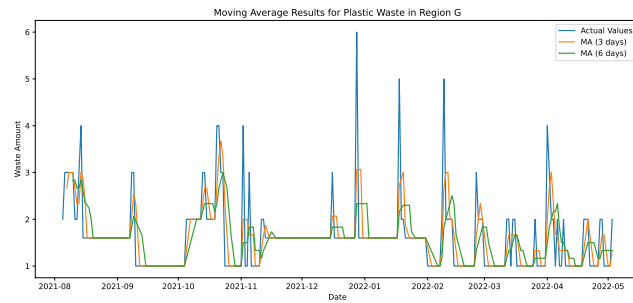


Figure 66. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region G

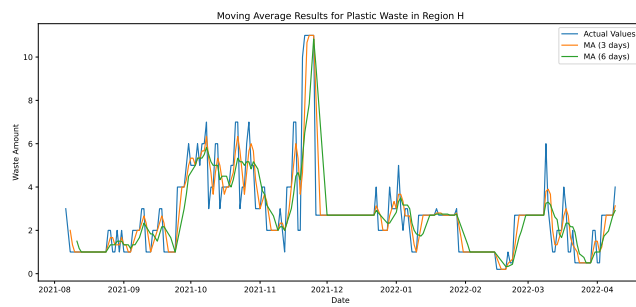


Figure 67. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region H

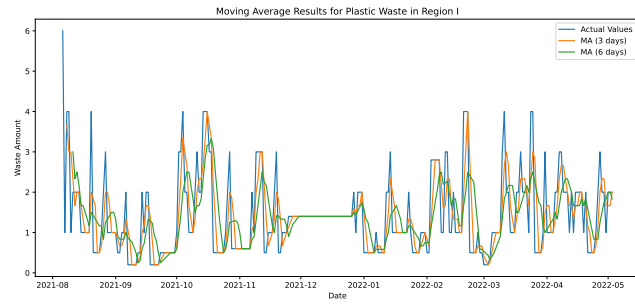


Figure 68. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region I

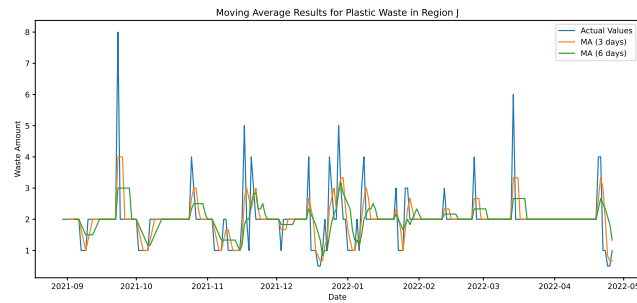


Figure 69. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region J

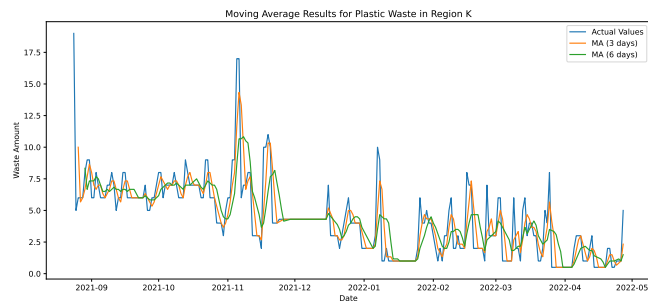


Figure 70. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region K

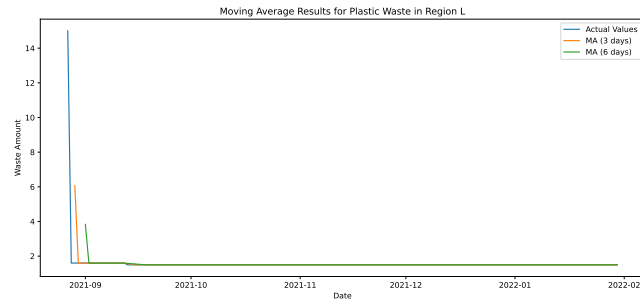


Figure 71. Moving Average Method Using 3 & 6 Days Period for Plastic Waste in Region L

Appendix C - Weighted Moving Average for Wastes in Regions

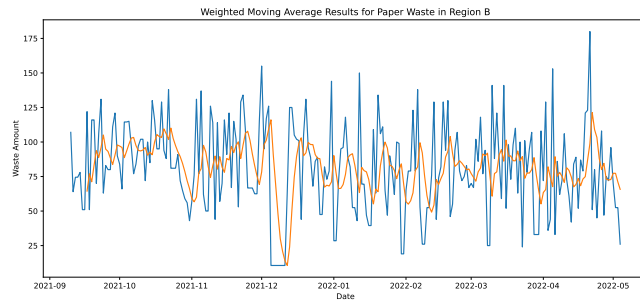


Figure 72. Weighted Moving Average for Paper Waste in Region B

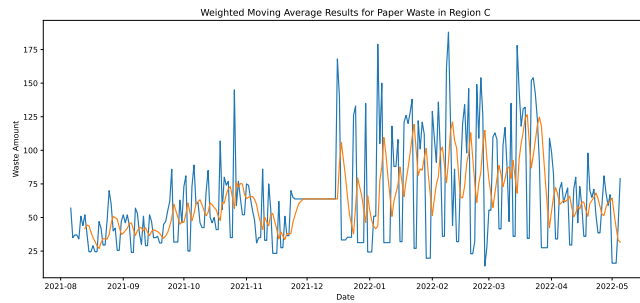


Figure 73. Weighted Moving Average for Paper Waste in Region C

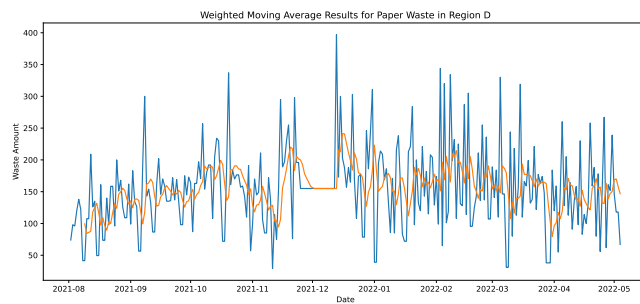


Figure 74. Weighted Moving Average for Paper Waste in Region D

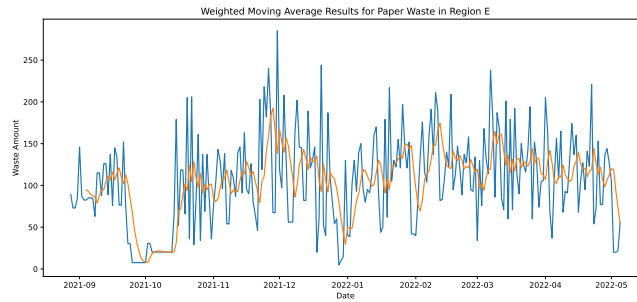


Figure 75. Weighted Moving Average for Paper Waste in Region E

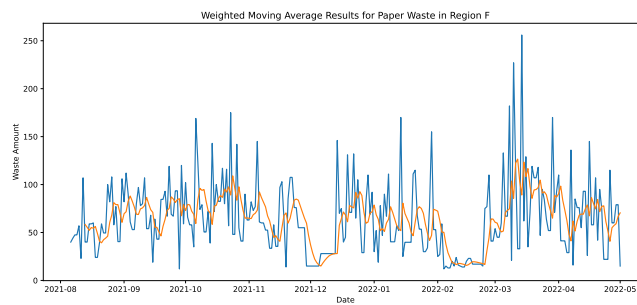


Figure 76. Weighted Moving Average for Paper Waste in Region F

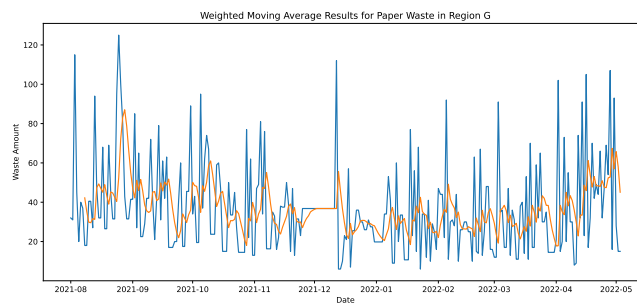


Figure 77. Weighted Moving Average for Paper Waste in Region G

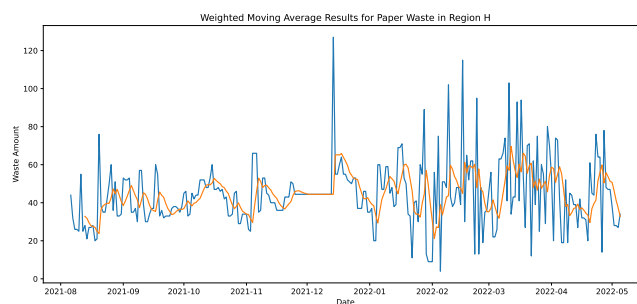


Figure 78. Weighted Moving Average for Paper Waste in Region H

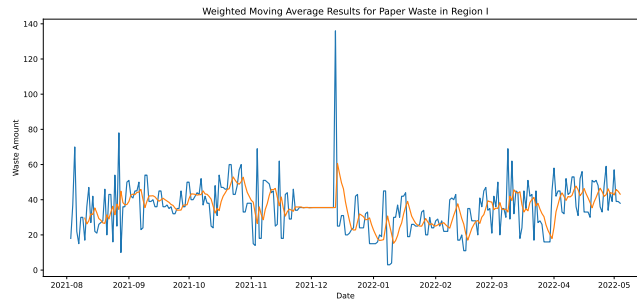


Figure 79. Weighted Moving Average for Paper Waste in Region I

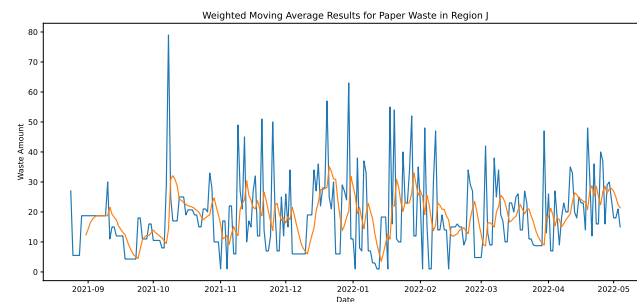


Figure 80. Weighted Moving Average for Paper Waste in Region J

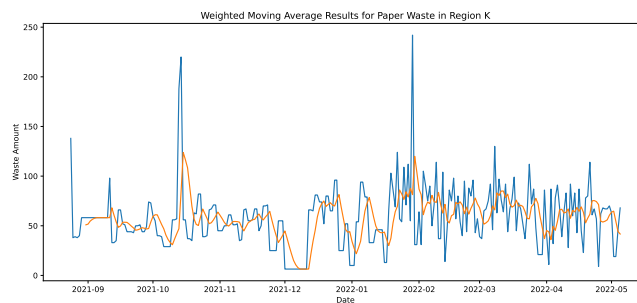


Figure 81. Weighted Moving Average for Paper Waste in Region K

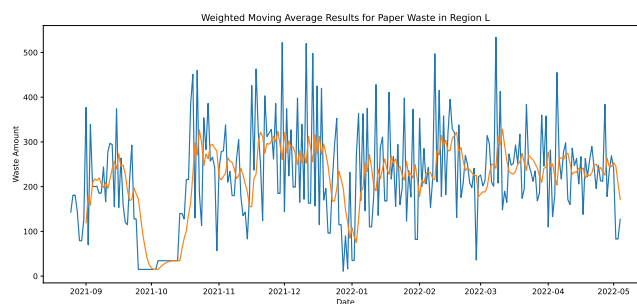


Figure 82. Weighted Moving Average for Paper Waste in Region L

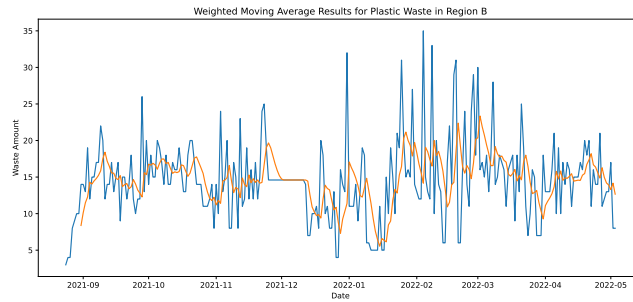


Figure 83. Weighted Moving Average for Plastic Waste in Region B

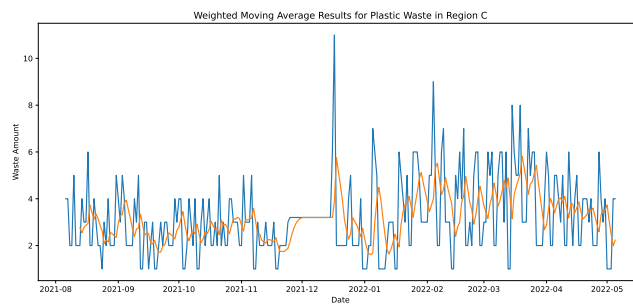


Figure 84. Weighted Moving Average for Plastic Waste in Region C

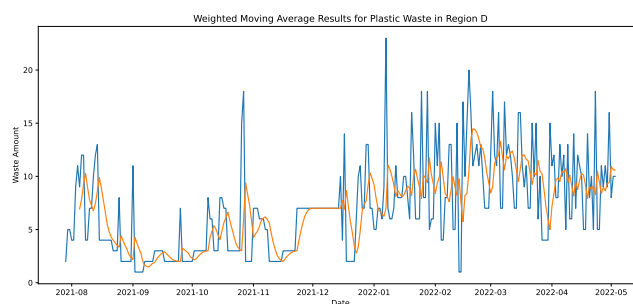


Figure 85. Weighted Moving Average for Plastic Waste in Region D

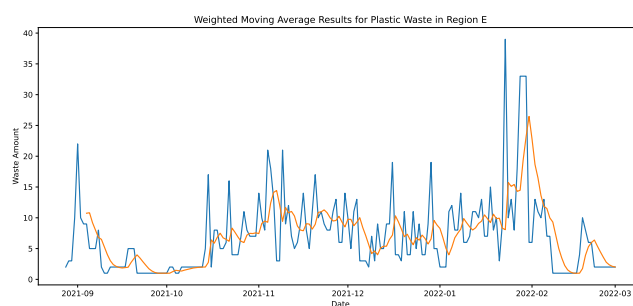


Figure 86. Weighted Moving Average for Plastic Waste in Region E

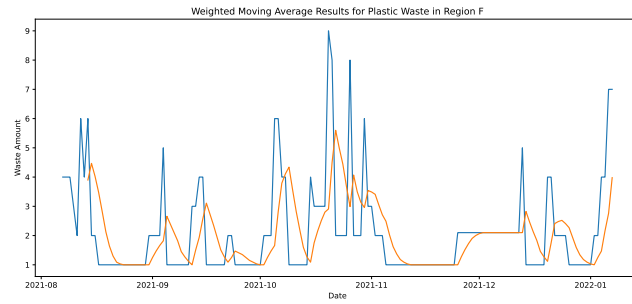


Figure 87. Weighted Moving Average for Plastic Waste in Region F

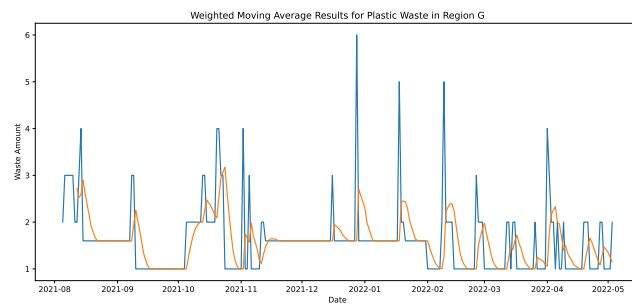


Figure 88. Weighted Moving Average for Plastic Waste in Region G

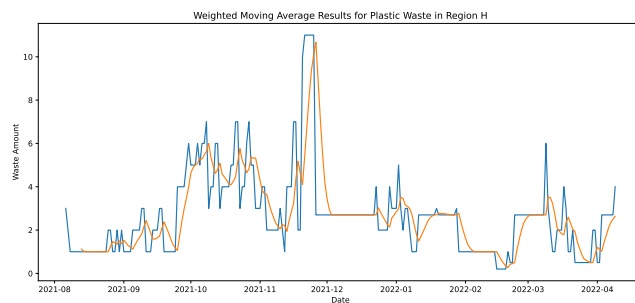


Figure 89. Weighted Moving Average for Plastic Waste in Region H

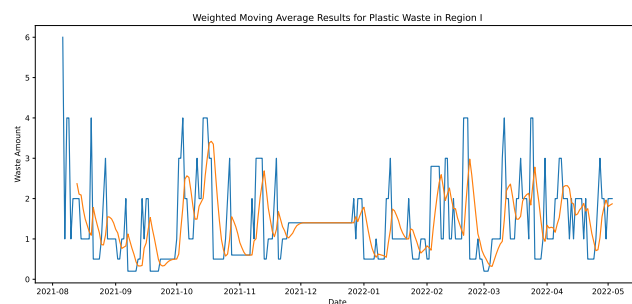


Figure 90. Weighted Moving Average for Plastic Waste in Region I

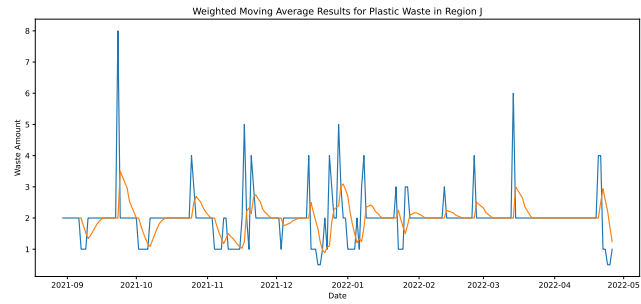


Figure 91. Weighted Moving Average for Plastic Waste in Region J

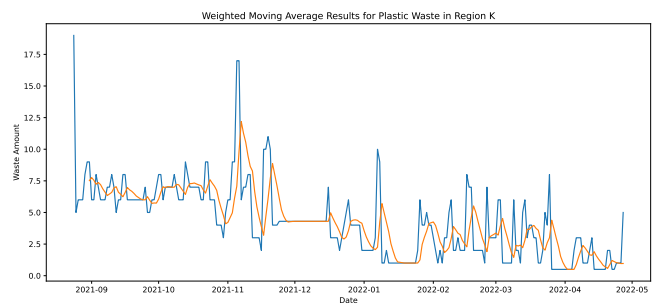


Figure 92. Weighted Moving Average for Plastic Waste in Region K

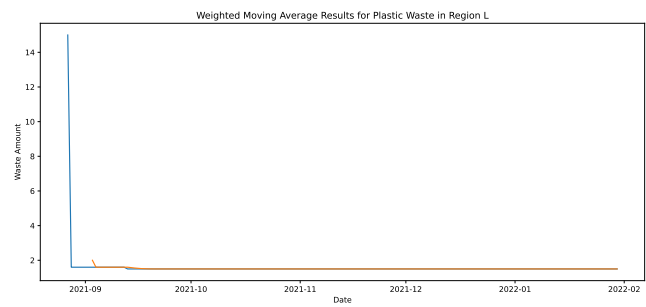


Figure 93. Weighted Moving Average for Plastic Waste in Region L

Appendix D - Exponential Smoothing Results for Wastes in Regions

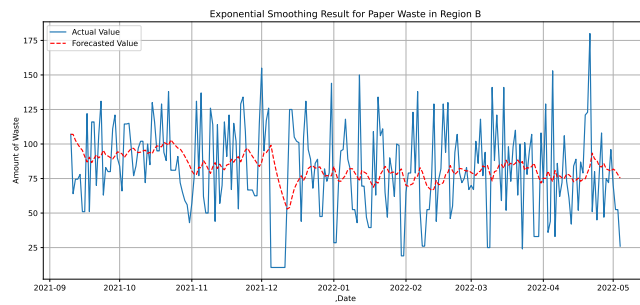


Figure 94. Exponential Smoothing Results for Paper Waste in Region B

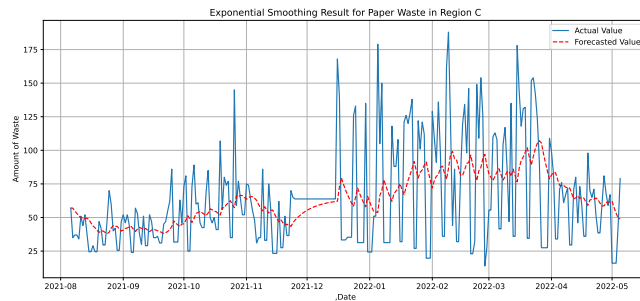


Figure 95. Exponential Smoothing Results for Paper Waste in Region C

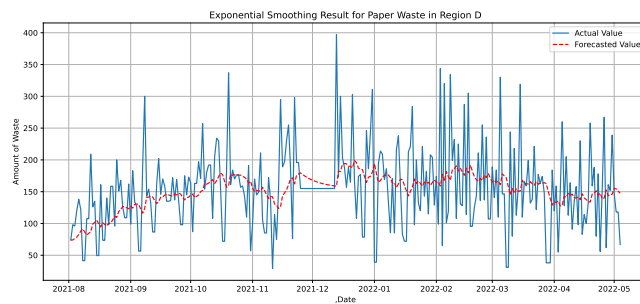


Figure 96. Exponential Smoothing Results for Paper Waste in Region D

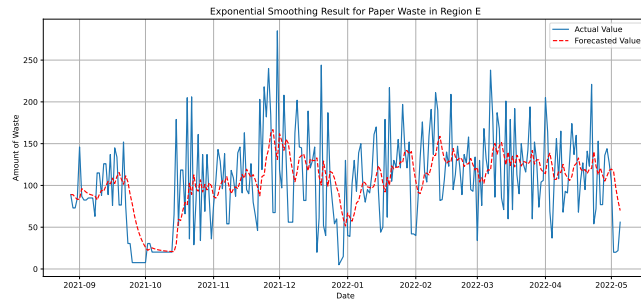


Figure 97. Exponential Smoothing Results for Paper Waste in Region E

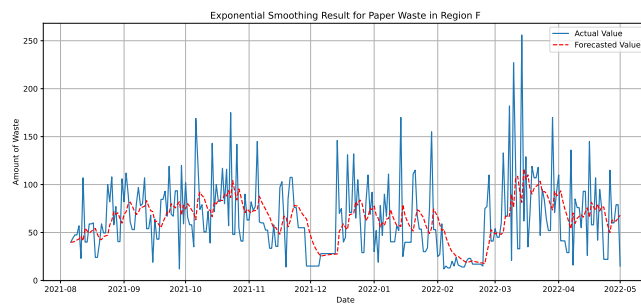


Figure 98. Exponential Smoothing Results for Paper Waste in Region F

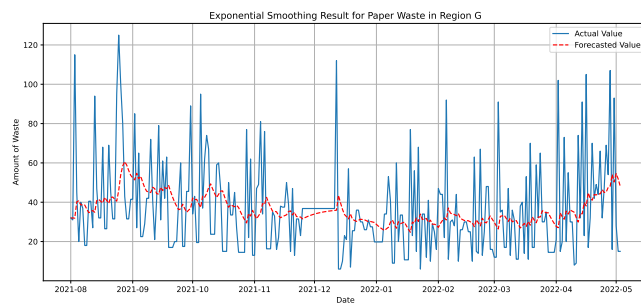


Figure 99. Exponential Smoothing Results for Paper Waste in Region G

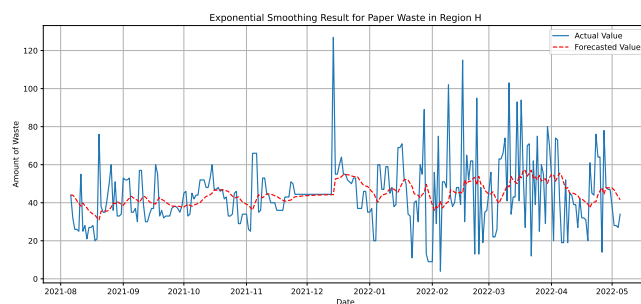


Figure 100. Exponential Smoothing Results for Paper Waste in Region H

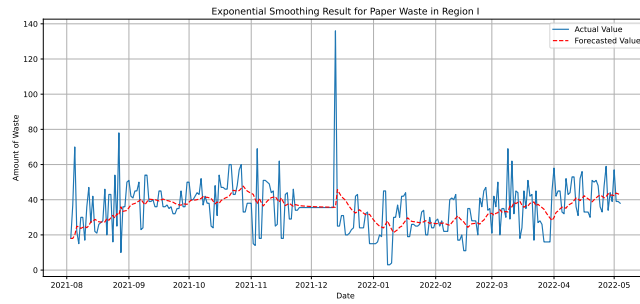


Figure 101. Exponential Smoothing Results for Paper Waste in Region I

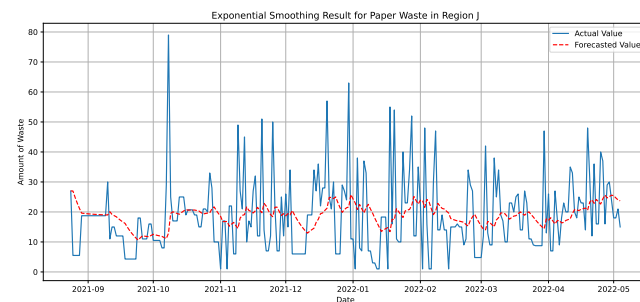


Figure 102. Exponential Smoothing Results for Paper Waste in Region J

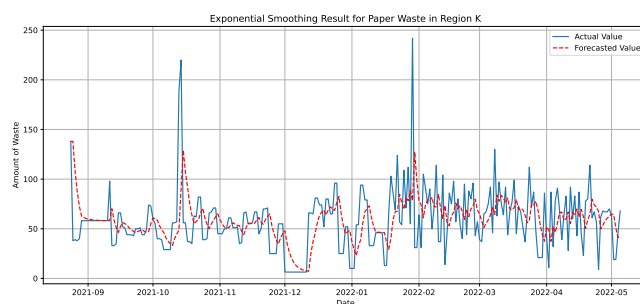


Figure 103. Exponential Smoothing Results for Paper Waste in Region K

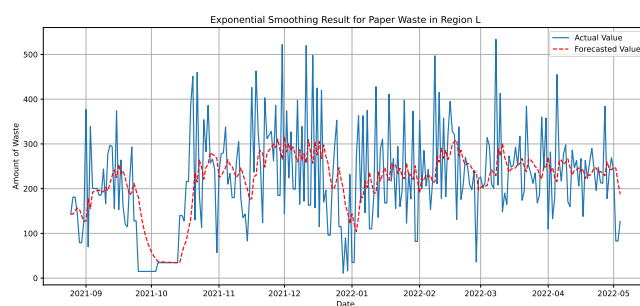


Figure 104. Exponential Smoothing Results for Paper Waste in Region L

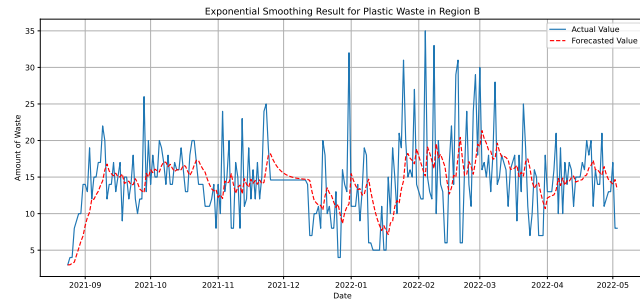


Figure 105. Exponential Smoothing Results for Plastic Waste in Region B

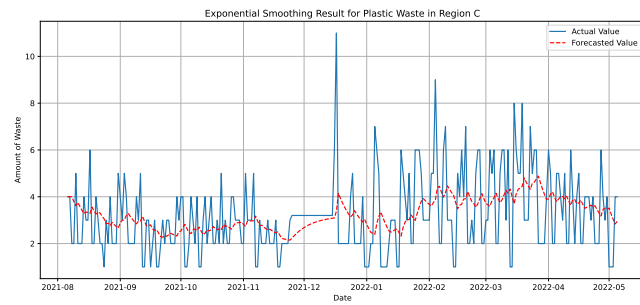


Figure 106. Exponential Smoothing Results for Plastic Waste in Region C

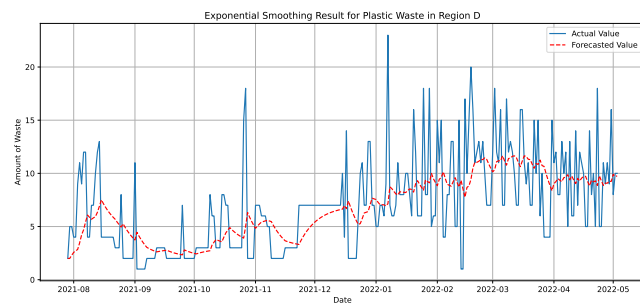


Figure 107. Exponential Smoothing Results for Plastic Waste in Region D

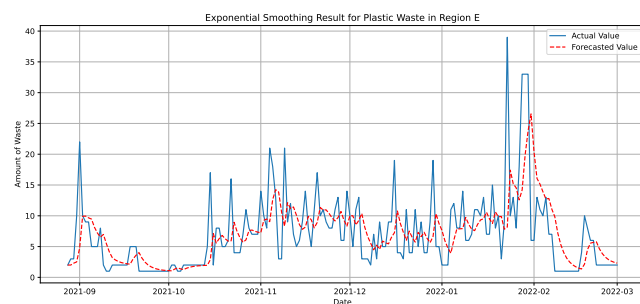


Figure 108. Exponential Smoothing Results for Plastic Waste in Region E

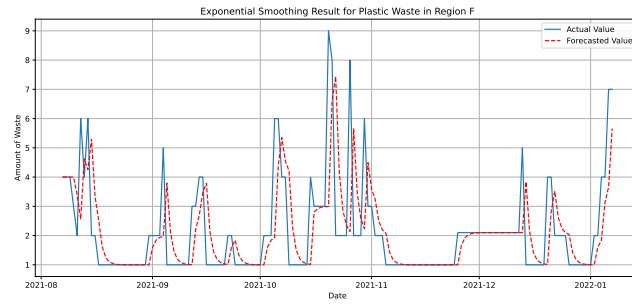


Figure 109. Exponential Smoothing Results for Plastic Waste in Region F

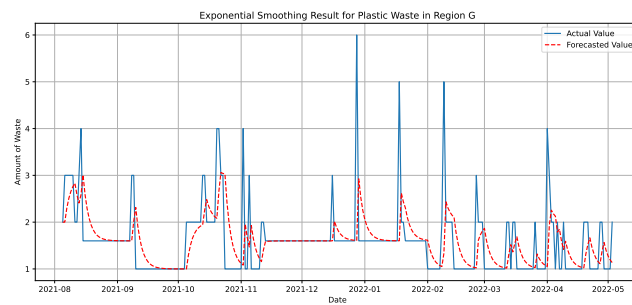


Figure 110. Exponential Smoothing Results for Plastic Waste in Region G

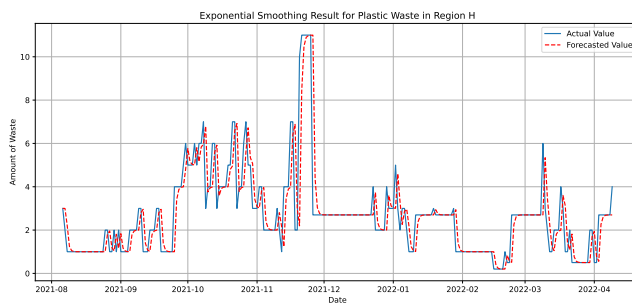


Figure 111. Exponential Smoothing Results for Plastic Waste in Region H

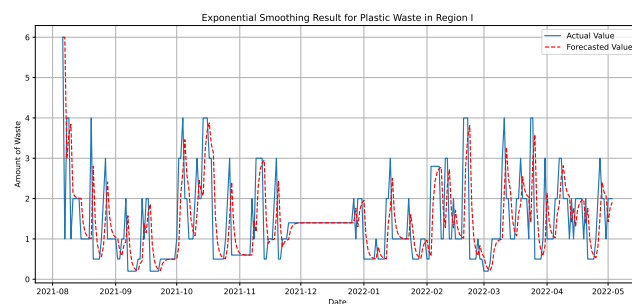


Figure 112. Exponential Smoothing Results for Plastic Waste in Region I

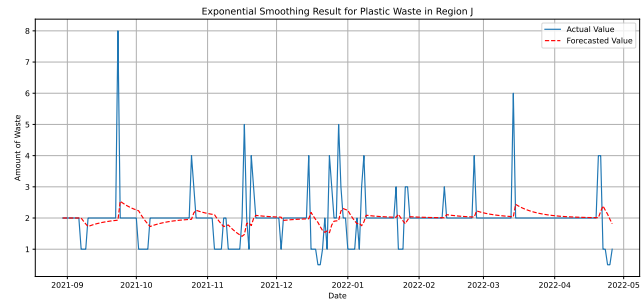


Figure 113. Exponential Smoothing Results for Plastic Waste in Region J

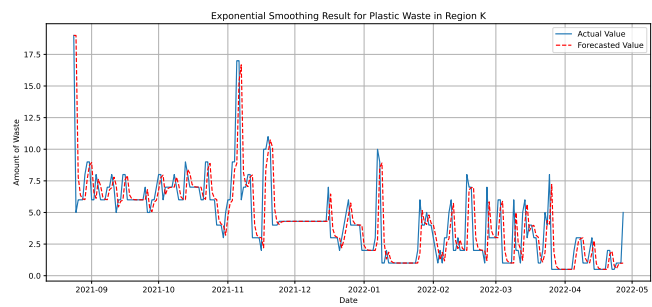


Figure 114. Exponential Smoothing Results for Plastic Waste in Region K

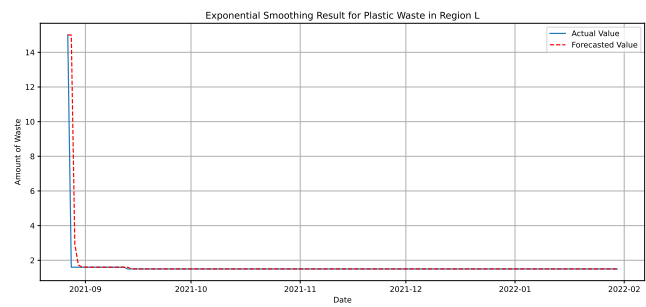


Figure 115. Exponential Smoothing Results for Plastic Waste in Region L

Appendix E - Holt-Winters Results for Wastes in Regions

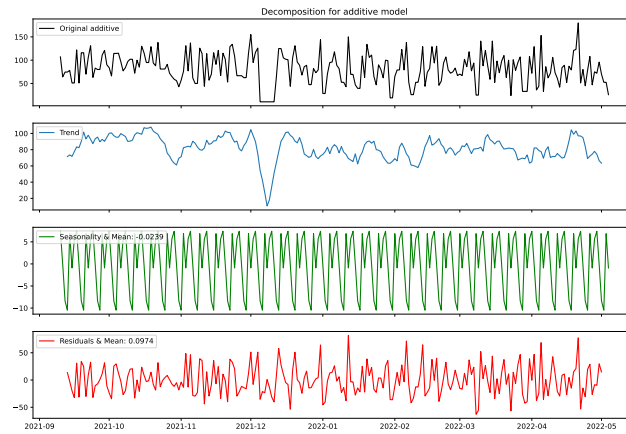


Figure 116. Decomposition of Time Series Model for Paper Waste in Region B

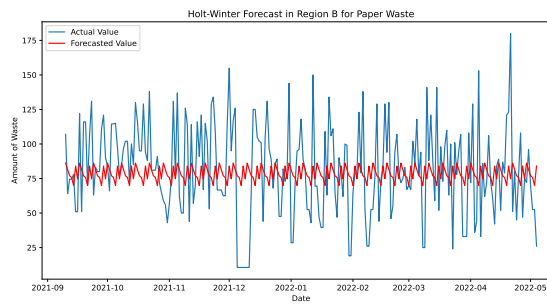


Figure 117. Holt-Winters Results for Paper Waste in Region B

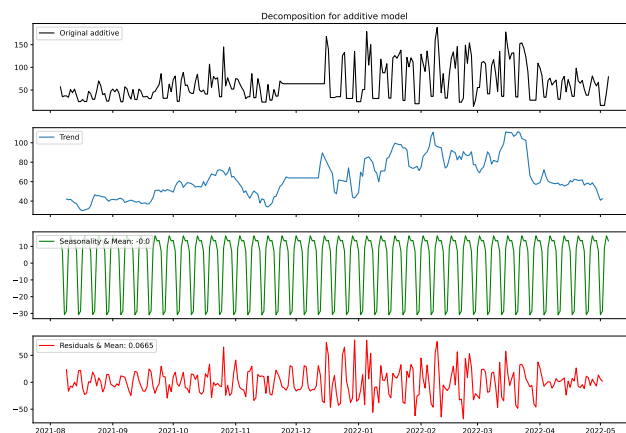


Figure 118. Decomposition of Time Series Model for Paper Waste in Region C

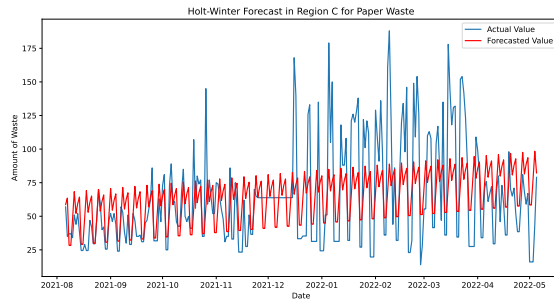


Figure 119. Holt-Winters Results for Paper Waste in Region C

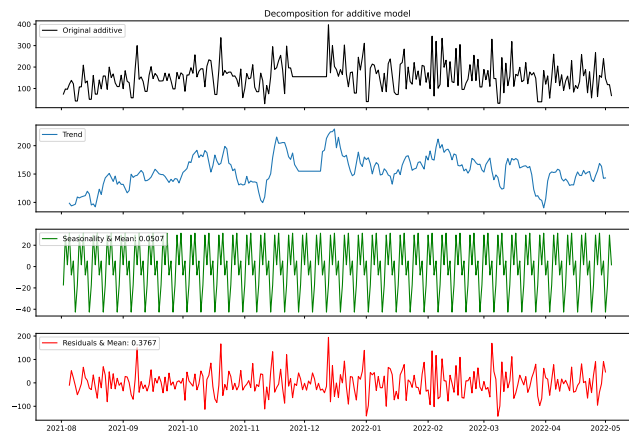


Figure 120. Decomposition of Time Series Model for Paper Waste in Region D

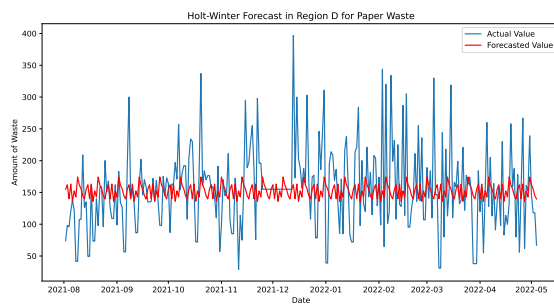


Figure 121. Holt-Winters Results for Paper Waste in Region D

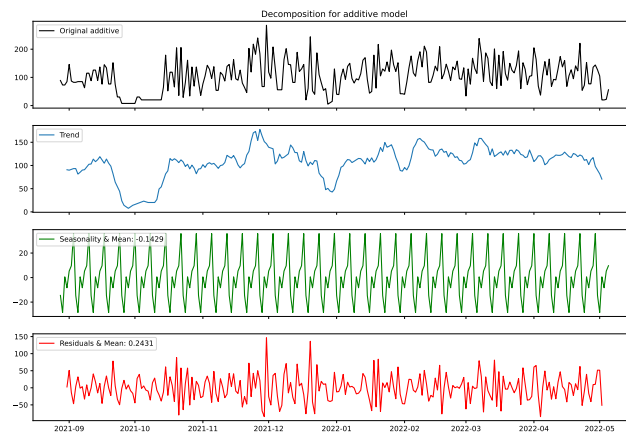


Figure 122. Decomposition of Time Series Model for Paper Waste in Region E

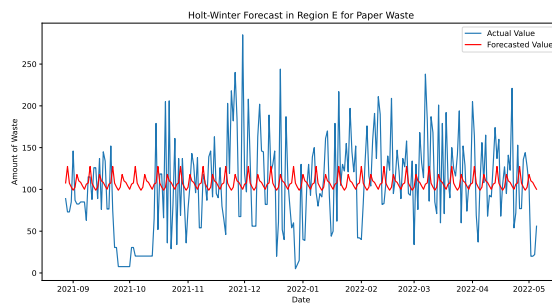


Figure 123. Holt-Winters Results for Paper Waste in Region E

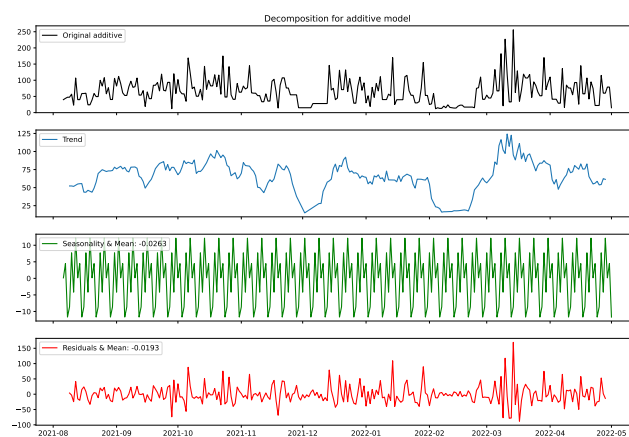


Figure 124. Decomposition of Time Series Model for Paper Waste in Region F

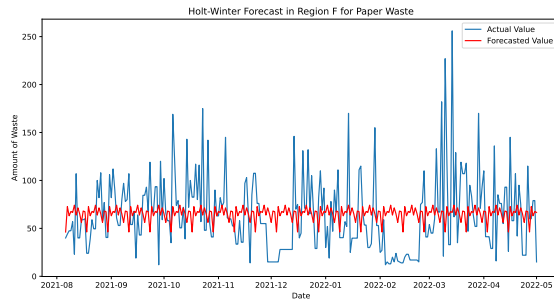


Figure 125. Holt-Winters Results for Paper Waste in Region F

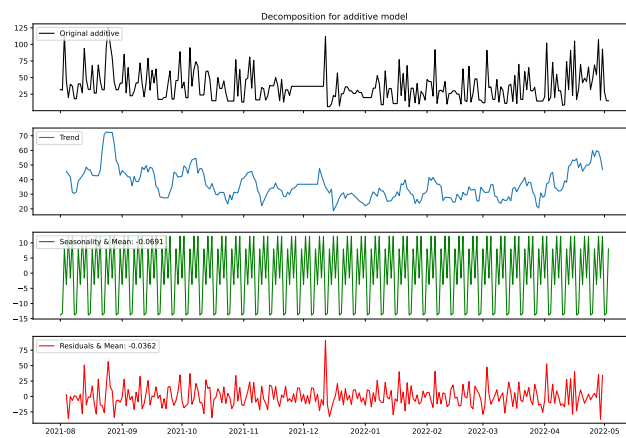


Figure 126. Decomposition of Time Series Model for Paper Waste in Region G

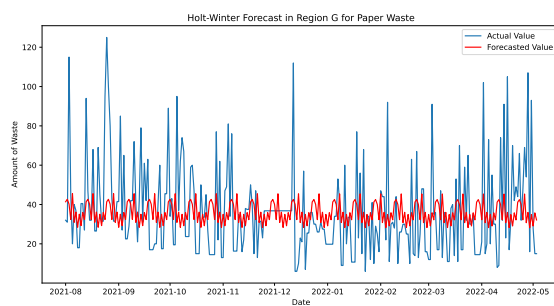


Figure 127. Holt-Winters Results for Paper Waste in Region G

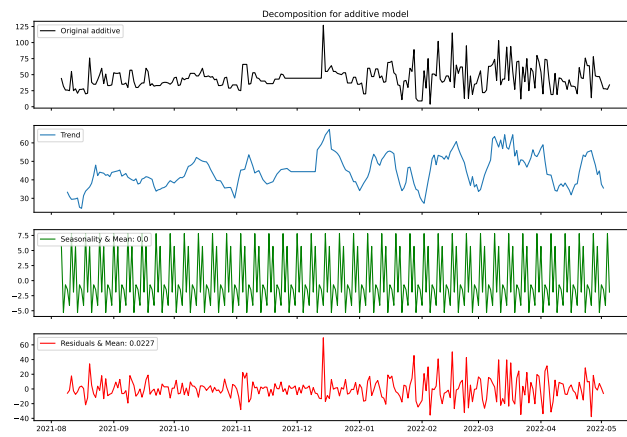


Figure 128. Decomposition of Time Series Model for Paper Waste in Region H

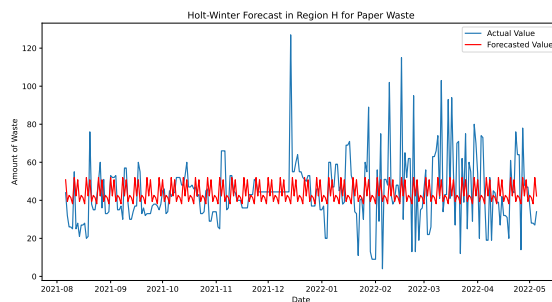


Figure 129. Holt-Winters Results for Paper Waste in Region H

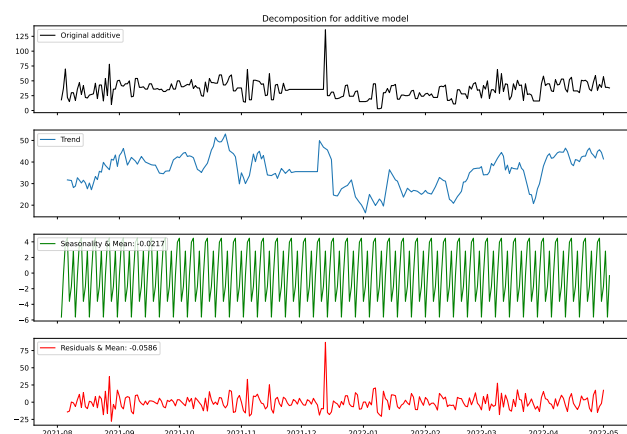


Figure 130. Decomposition of Time Series Model for Paper Waste in Region I

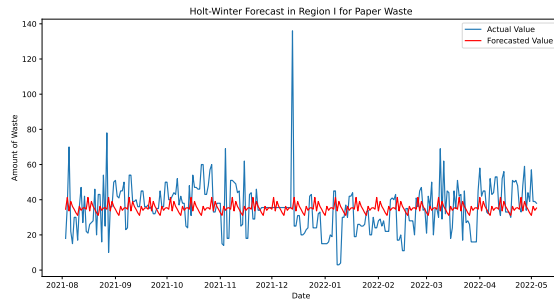


Figure 131. Holt-Winters Results for Paper Waste in Region I

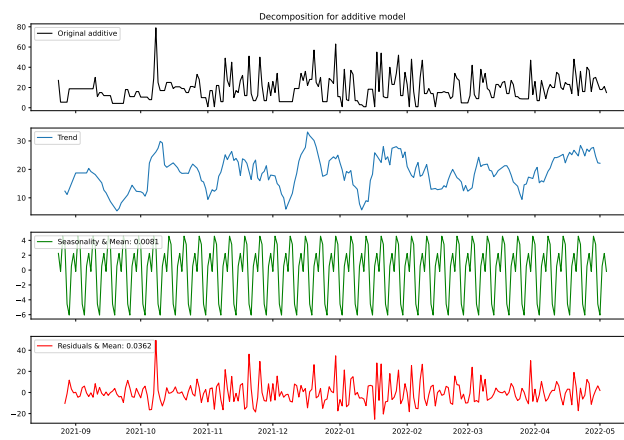


Figure 132. Decomposition of Time Series Model for Paper Waste in Region J

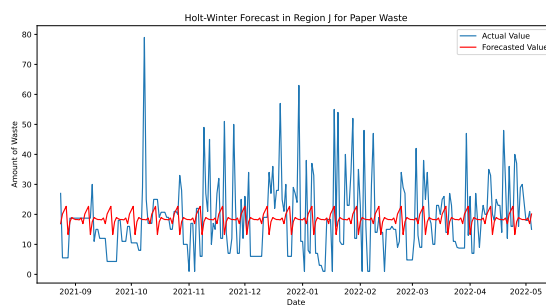


Figure 133. Holt-Winters Results for Paper Waste in Region J

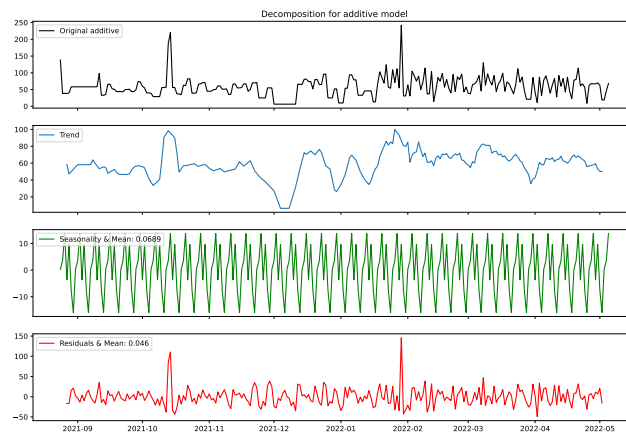


Figure 134. Decomposition of Time Series Model for Paper Waste in Region K

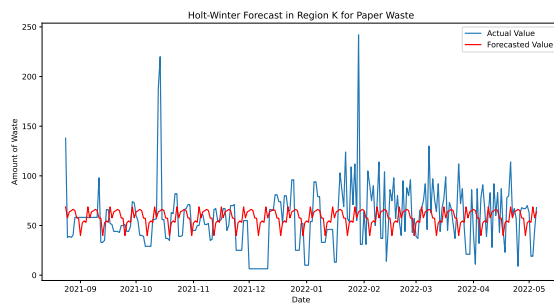


Figure 135. Holt-Winters Results for Paper Waste in Region K

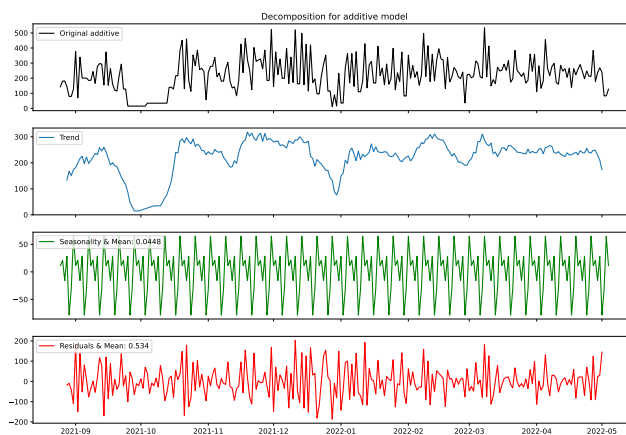


Figure 136. Decomposition of Time Series Model for Paper Waste in Region L

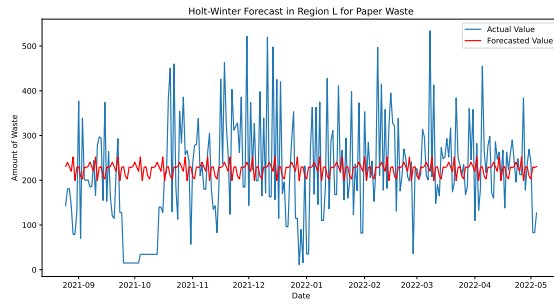


Figure 137. Holt-Winters Results for Paper Waste in Region L

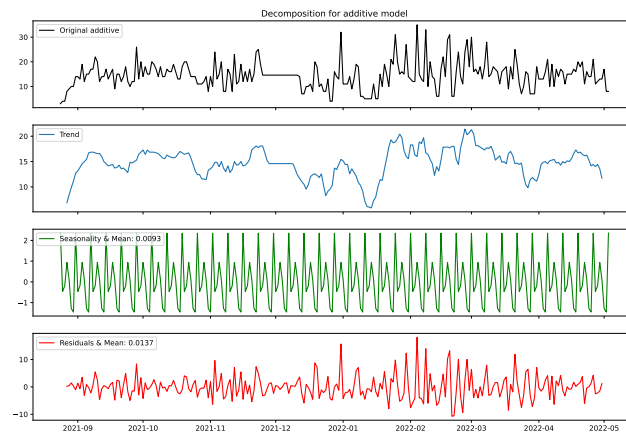


Figure 138. Decomposition of Time Series Model for Plastic Waste in Region B

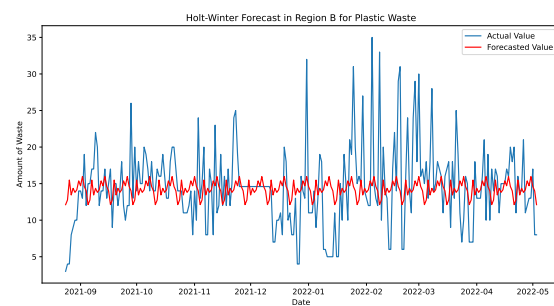


Figure 139. Holt-Winters Results for Plastic Waste in Region B

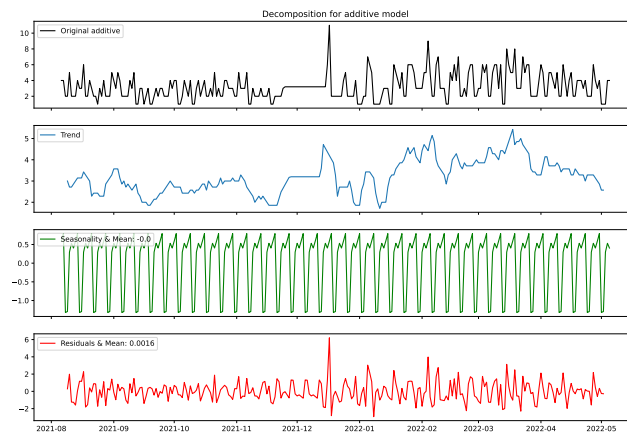


Figure 140. Decomposition of Time Series Model for Plastic Waste in Region C

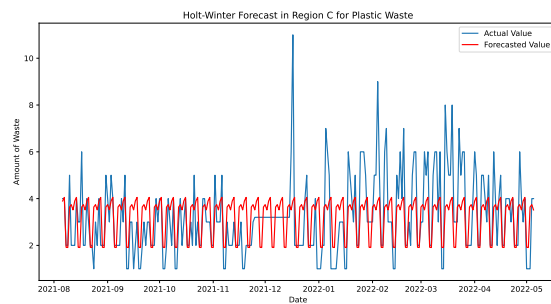


Figure 141. Holt-Winters Results for Plastic Waste in Region C

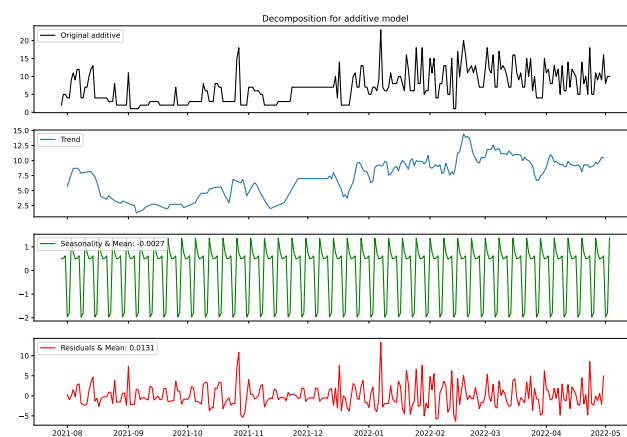


Figure 142. Decomposition of Time Series Model for Plastic Waste in Region D

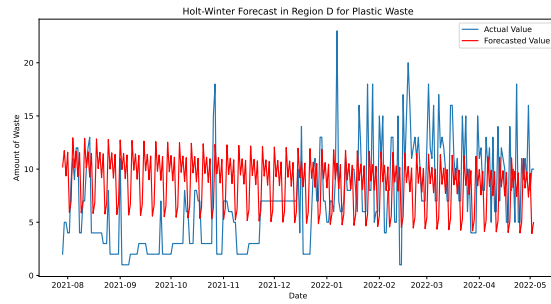


Figure 143. Holt-Winters Results for Plastic Waste in Region D

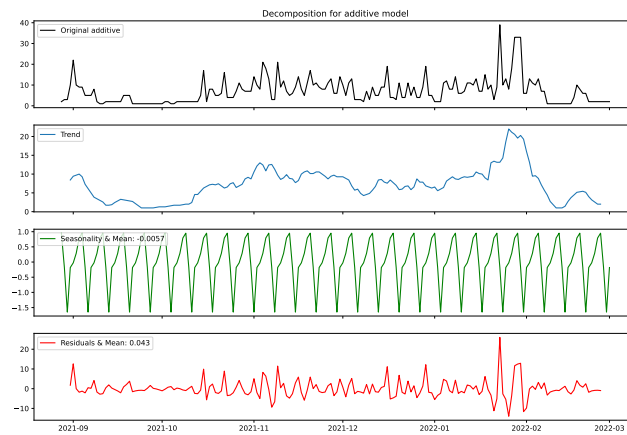


Figure 144. Decomposition of Time Series Model for Plastic Waste in Region E

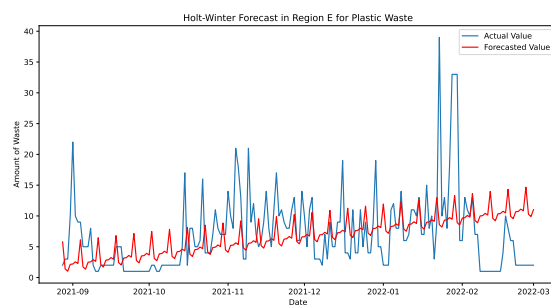


Figure 145. Holt-Winters Results for Plastic Waste in Region E

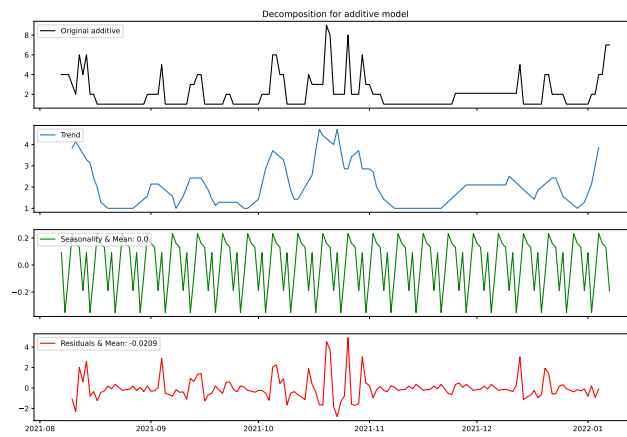


Figure 146. Decomposition of Time Series Model for Plastic Waste in Region F

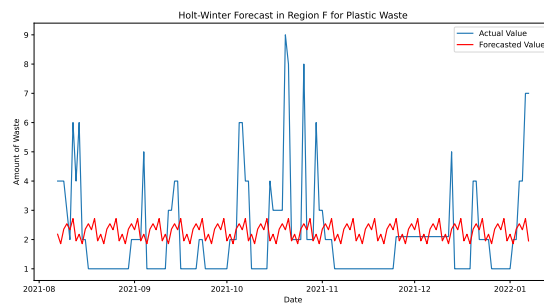


Figure 147. Holt-Winters Results for Plastic Waste in Region F

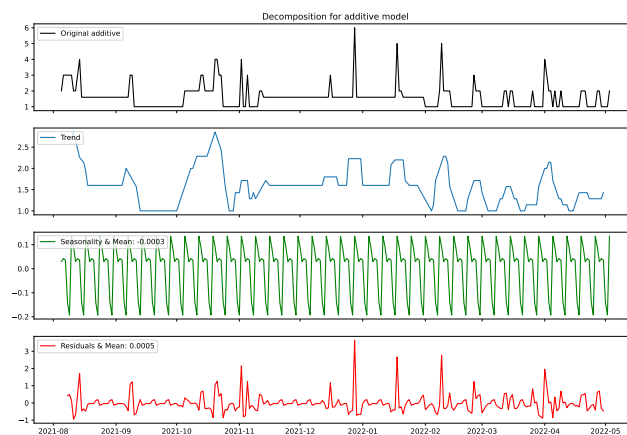


Figure 148. Decomposition of Time Series Model for Plastic Waste in Region G

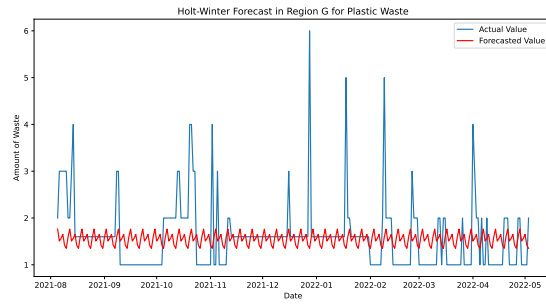


Figure 149. Holt-Winters Results for Plastic Waste in Region G

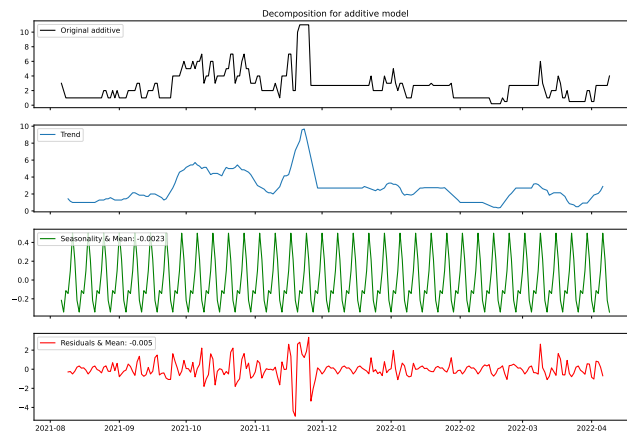


Figure 150. Decomposition of Time Series Model for Plastic Waste in Region H

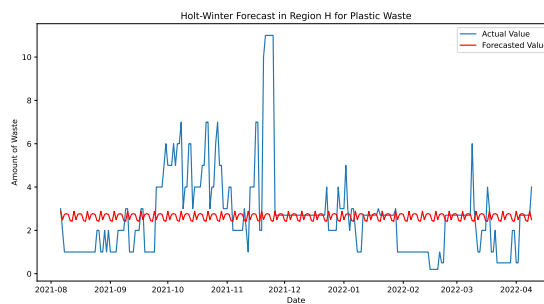


Figure 151. Holt-Winters Results for Plastic Waste in Region H

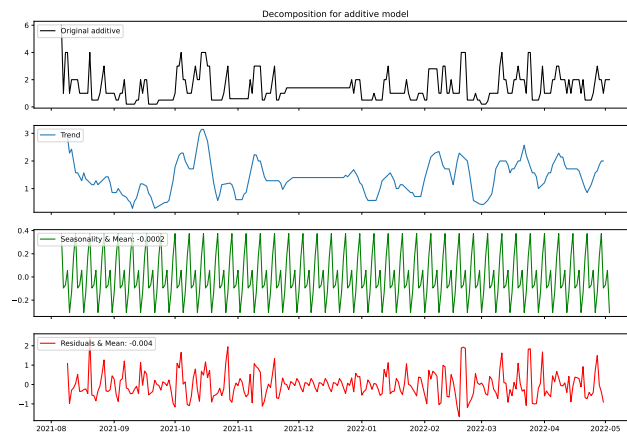


Figure 152. Decomposition of Time Series Model for Plastic Waste in Region I

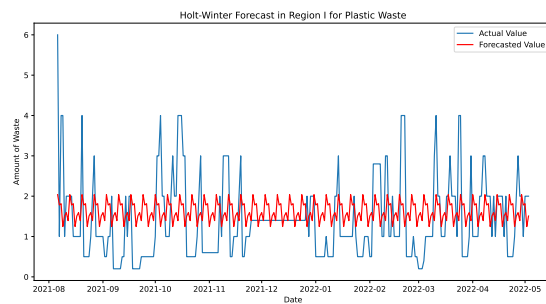


Figure 153. Holt-Winters Results for Plastic Waste in Region I

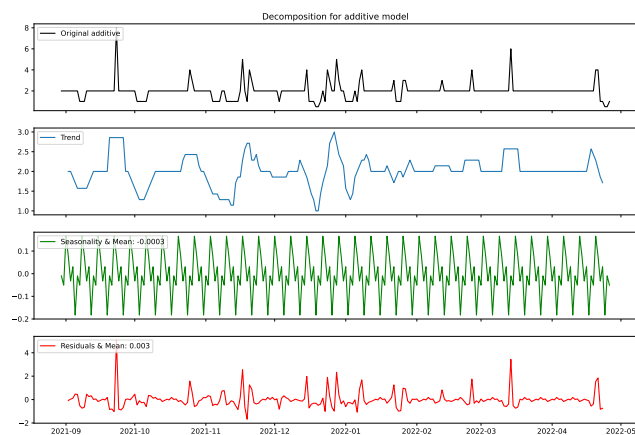


Figure 154. Decomposition of Time Series Model for Plastic Waste in Region J

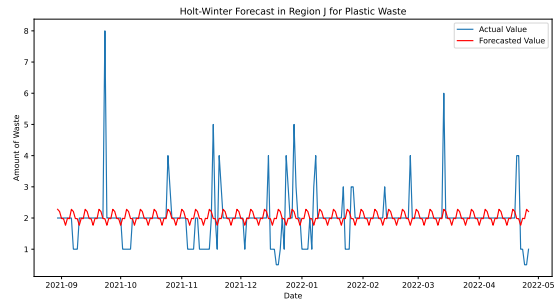


Figure 155. Holt-Winters Results for Plastic Waste in Region J

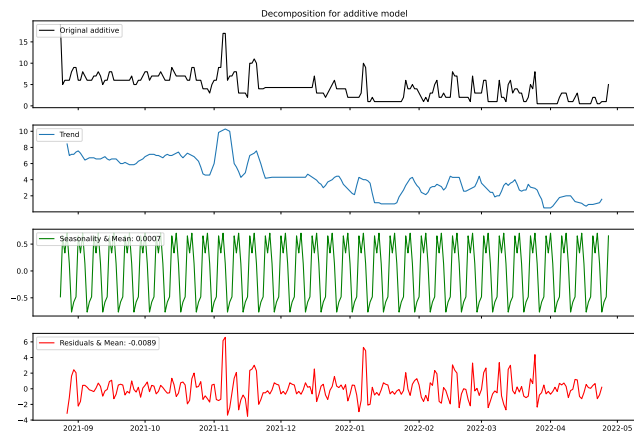


Figure 156. Decomposition of Time Series Model for Plastic Waste in Region K

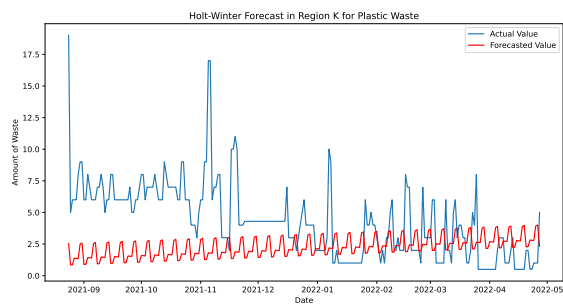


Figure 157. Holt-Winters Results for Plastic Waste in Region K

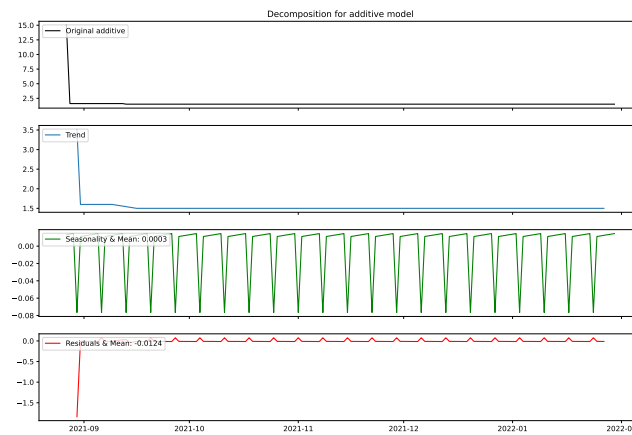


Figure 158. Decomposition of Time Series Model for Plastic Waste in Region L

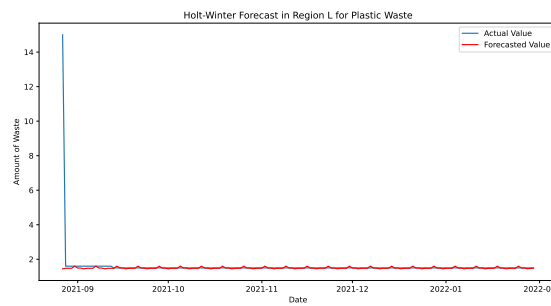


Figure 159. Holt-Winters Results for Plastic Waste in Region L

Appendix F - Graphs of Actual and Forecasted Values using XGBoost

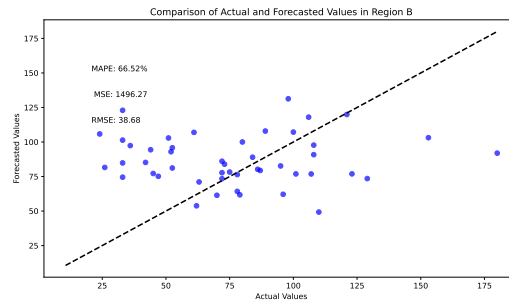


Figure 160. Graphs of Actual and Forecasted Values for Paper Waste in Region B

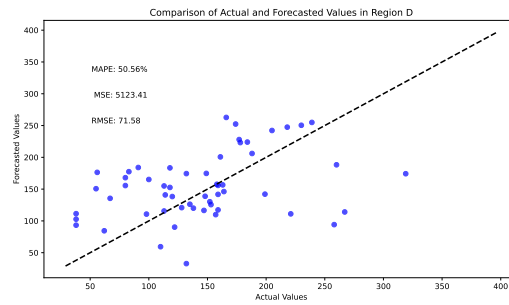


Figure 161. Graphs of Actual and Forecasted Values for Paper Waste in Region D

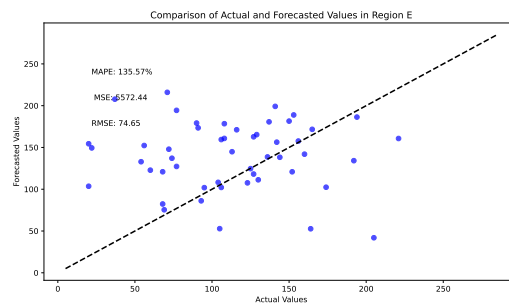


Figure 162. Graphs of Actual and Forecasted Values for Paper Waste in Region E

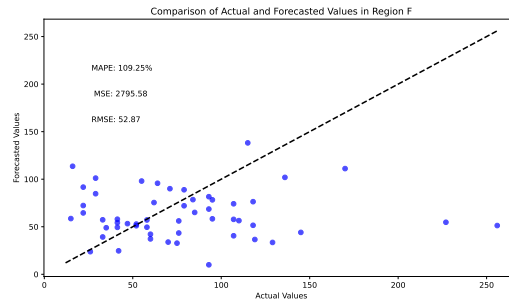


Figure 163. Graphs of Actual and Forecasted Values for Paper Waste in Region F

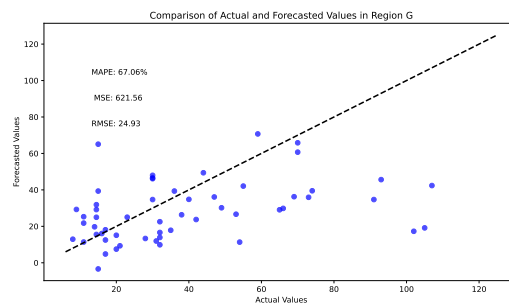


Figure 164. Graphs of Actual and Forecasted Values for Paper Waste in Region G

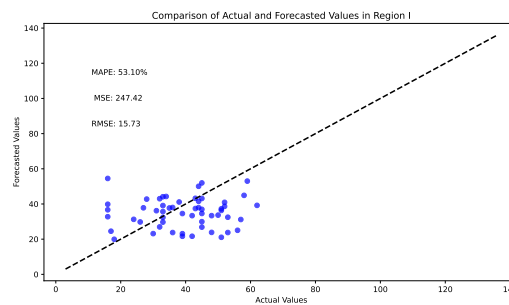


Figure 165. Graphs of Actual and Forecasted Values for Paper Waste in Region I

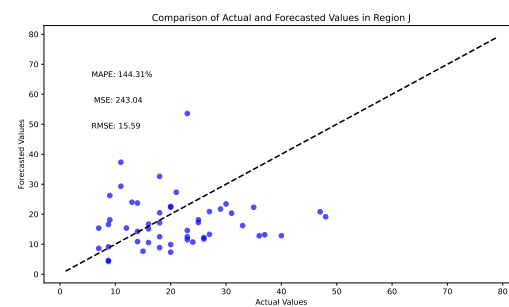


Figure 166. Graphs of Actual and Forecasted Values for Paper Waste in Region J

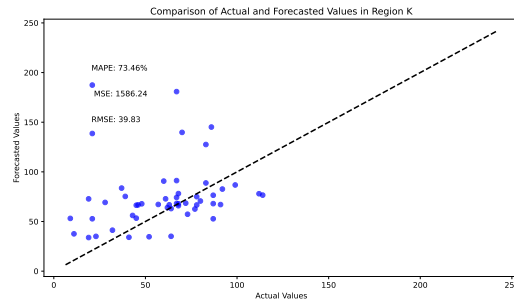


Figure 167. Graphs of Actual and Forecasted Values for Paper Waste in Region K

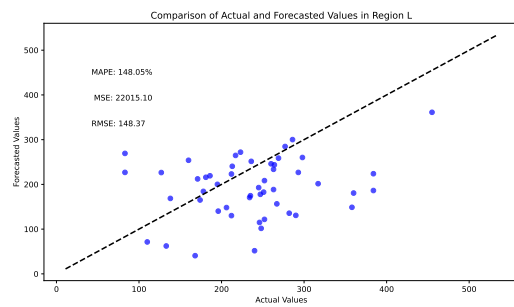


Figure 168. Graphs of Actual and Forecasted Values for Paper Waste in Region L

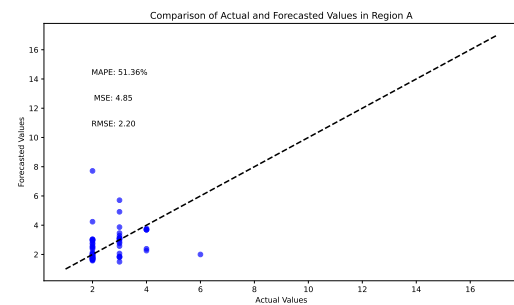


Figure 169. Graphs of Actual and Forecasted Values for Plastic Waste in Region A

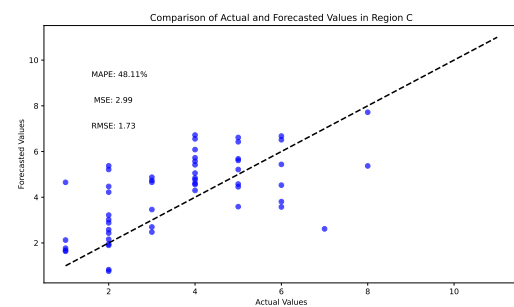


Figure 170. Graphs of Actual and Forecasted Values for Plastic Waste in Region C

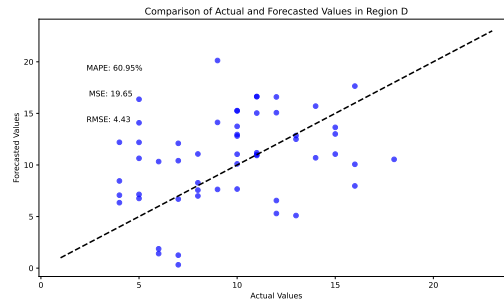


Figure 171. Graphs of Actual and Forecasted Values for Plastic Waste in Region D

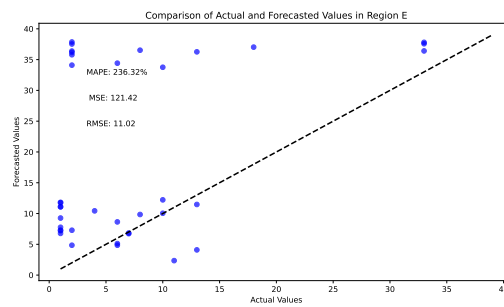


Figure 172. Graphs of Actual and Forecasted Values for Plastic Waste in Region E

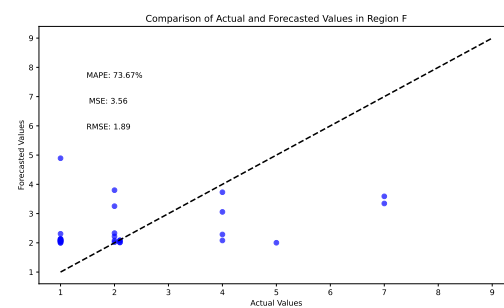


Figure 173. Graphs of Actual and Forecasted Values for Plastic Waste in Region F

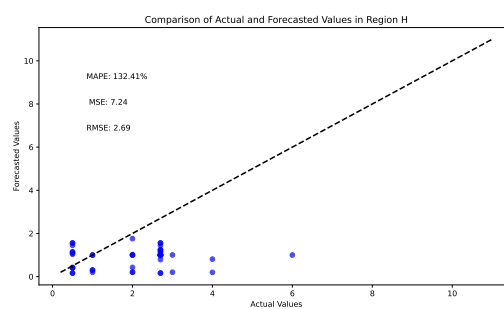


Figure 174. Graphs of Actual and Forecasted Values for Plastic Waste in Region H

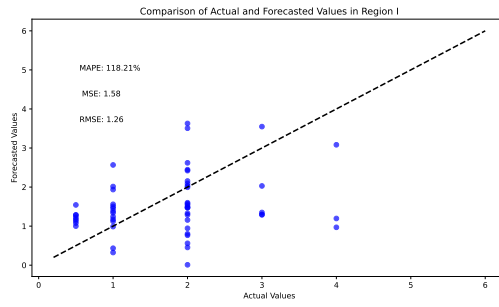


Figure 175. Graphs of Actual and Forecasted Values for Plastic Waste in Region I

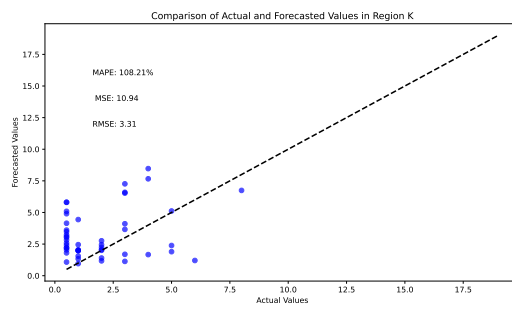


Figure 176. Graphs of Actual and Forecasted Values for Plastic Waste in Region K

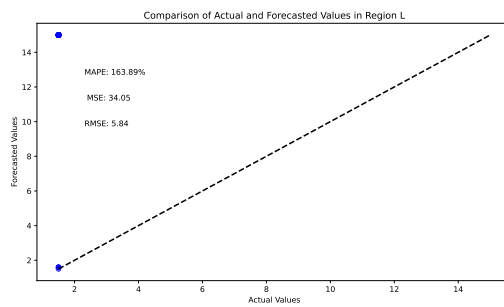


Figure 177. Graphs of Actual and Forecasted Values for Plastic Waste in Region L

Appendix G - Graphs of Actual and Forecasted Values using K-NN

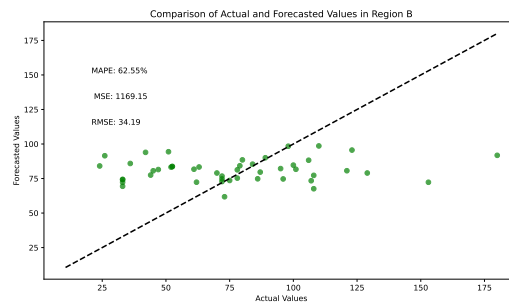


Figure 178. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region B

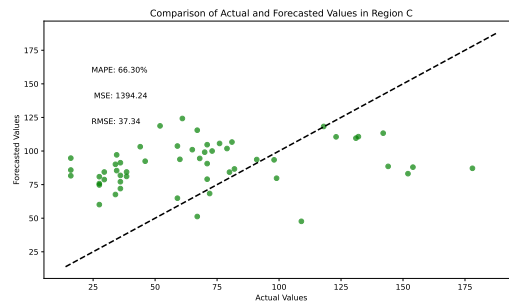


Figure 179. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region C

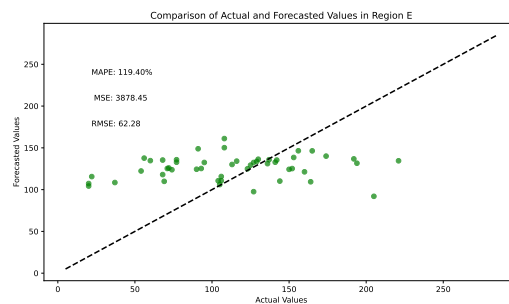


Figure 180. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region E

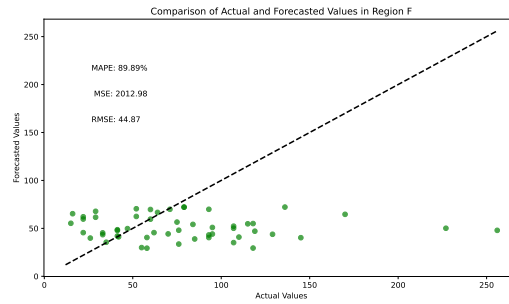


Figure 181. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region F

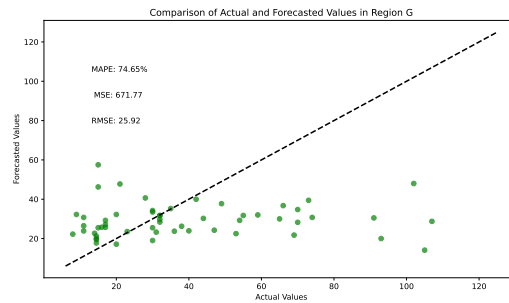


Figure 182. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region G

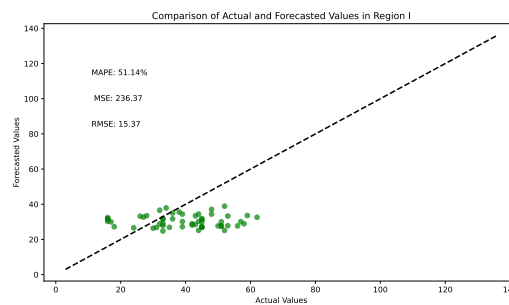


Figure 183. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region I

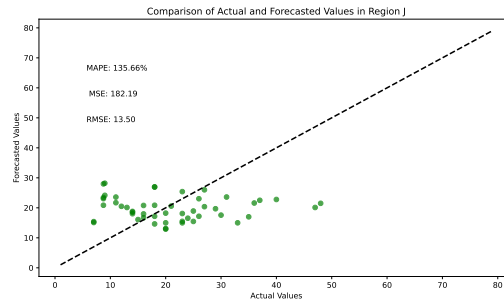


Figure 184. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region J

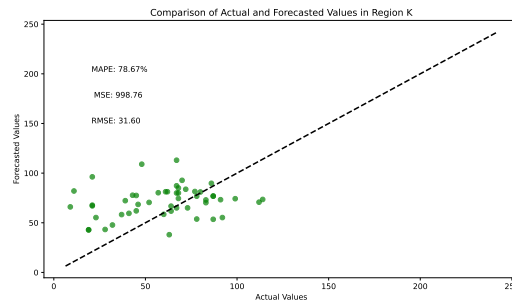


Figure 185. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region K

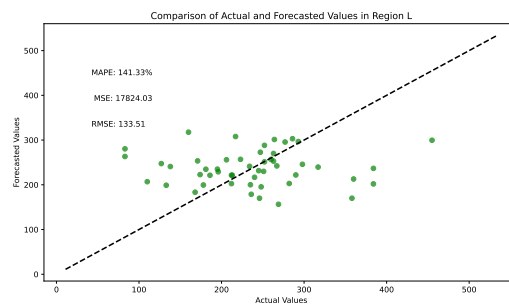


Figure 186. Graphs of Forecasted Data Corresponding to Actual Data of Paper Waste in Region L

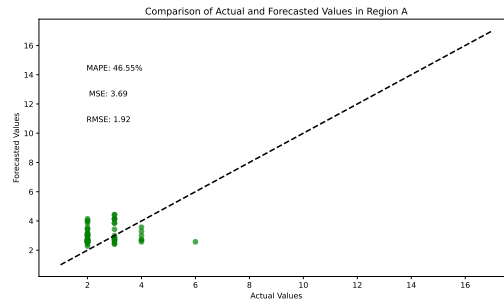


Figure 187. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region A

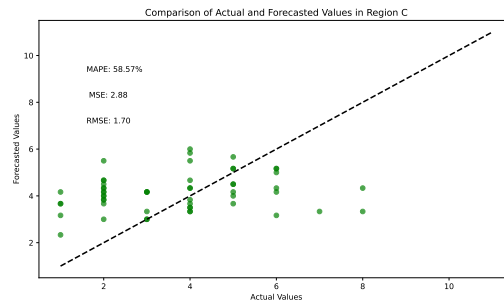


Figure 188. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region C

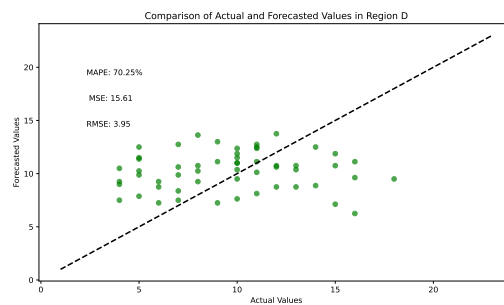


Figure 189. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region D

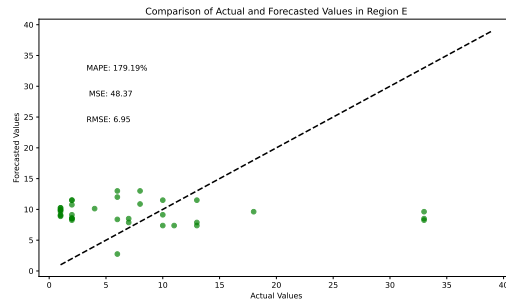


Figure 190. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region E

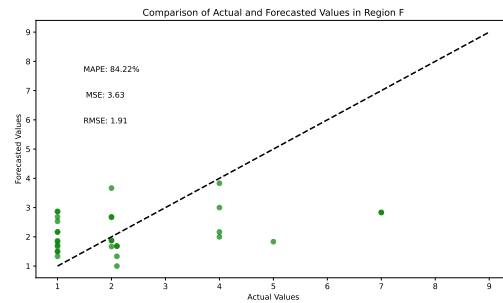


Figure 191. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region F

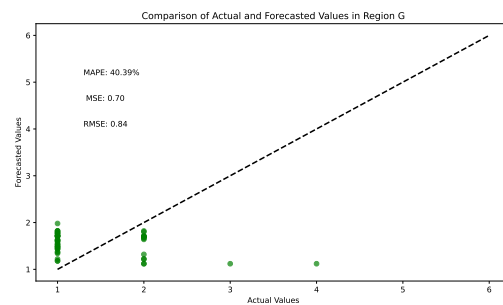


Figure 192. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region G

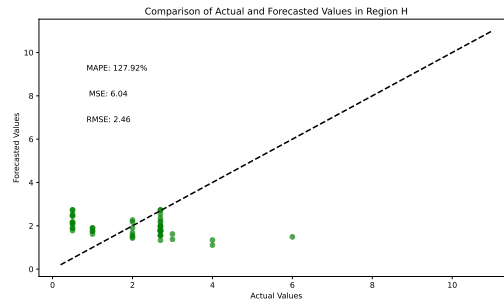


Figure 193. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region H

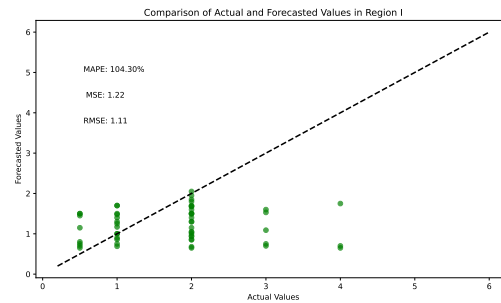


Figure 194. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region I

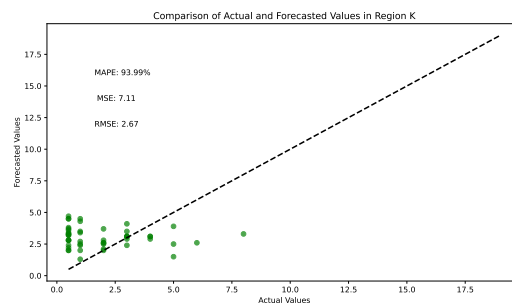


Figure 195. Graphs of Forecasted Data Corresponding to Actual Data of Plastic Waste in Region K

Appendix H - Graphs of Actual and Forecasted Values using ARIMA

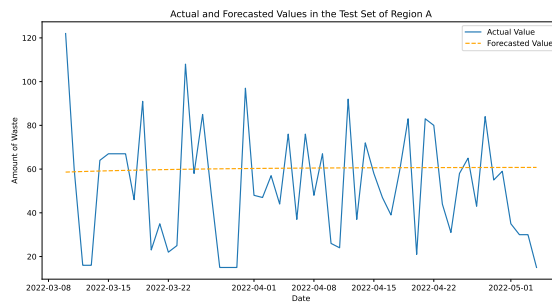


Figure 196. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region A

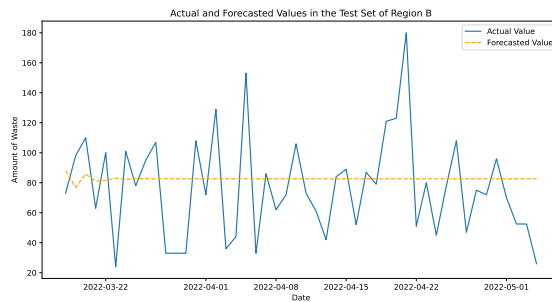


Figure 197. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region B

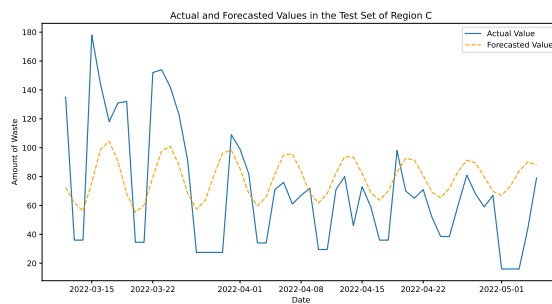


Figure 198. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region C

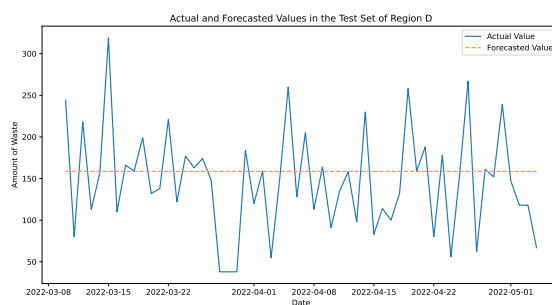


Figure 199. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region D

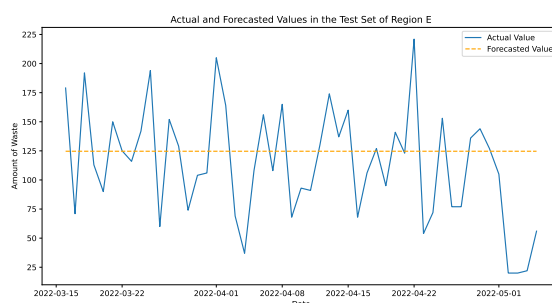


Figure 200. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region E

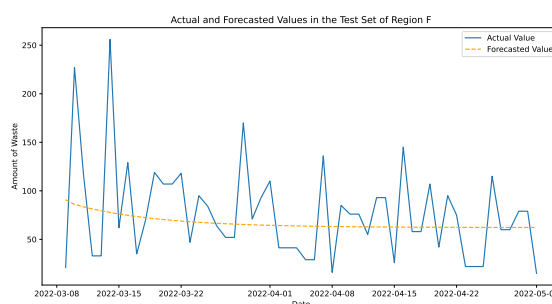


Figure 201. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region F

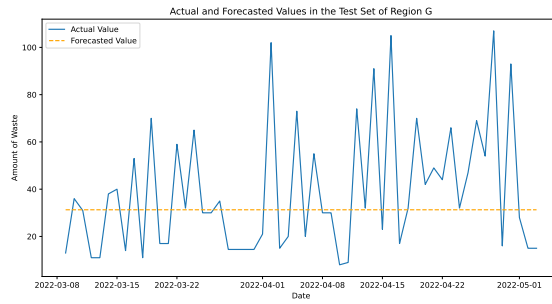


Figure 202. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region G

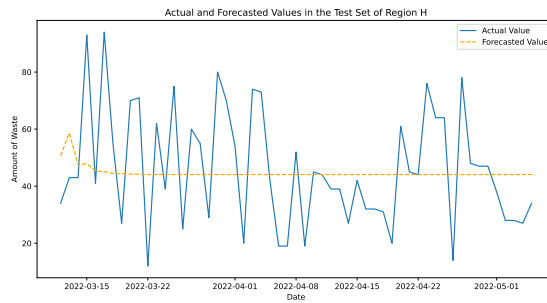


Figure 203. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region H

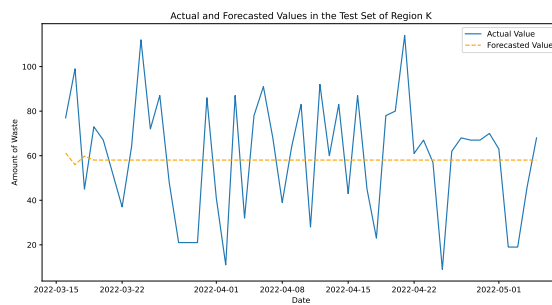


Figure 204. Graphs of Actual and Forecasted Values in the Testset of Paper Waste in Region K

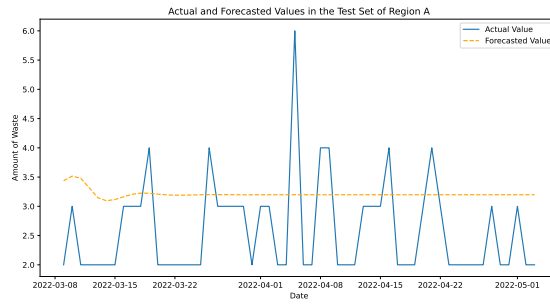


Figure 205. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region A

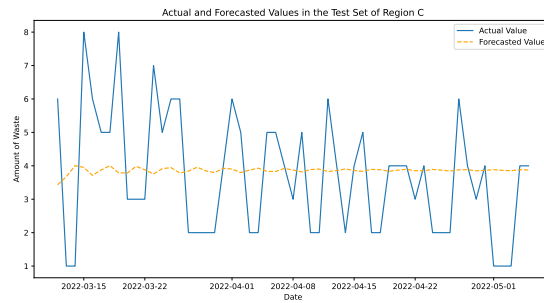


Figure 206. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region C

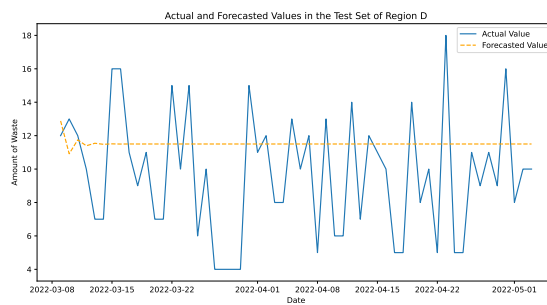


Figure 207. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region D

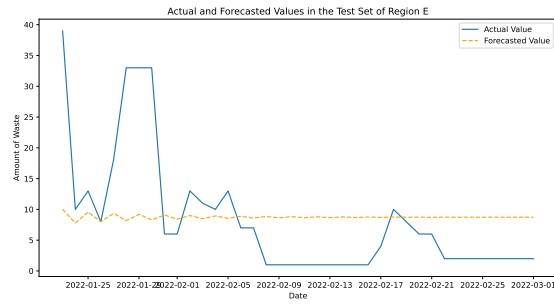


Figure 208. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region E

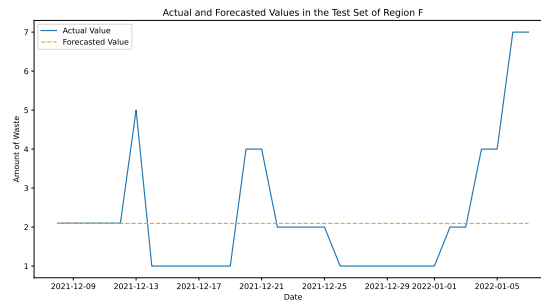


Figure 209. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region F

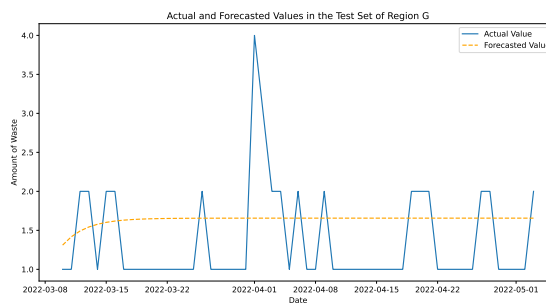


Figure 210. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region G

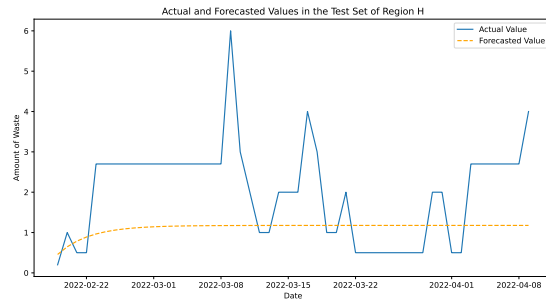


Figure 211. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region H

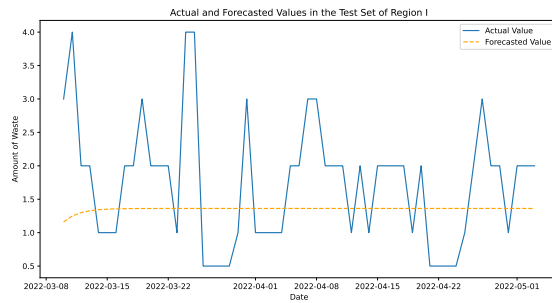


Figure 212. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region I

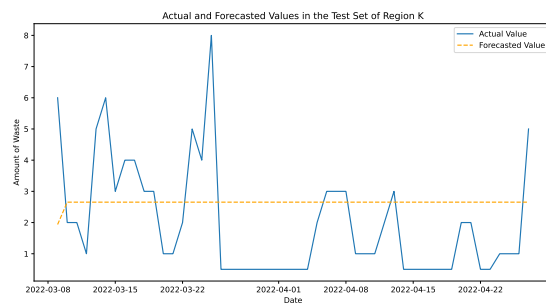


Figure 213. Graphs of Actual and Forecasted Values in the Testset of Plastic Waste in Region K