

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**DISCOVERING MARKET INSIGHTS FROM ONLINE PRODUCT REVIEWS
THROUGH SENTIMENT ANALYSIS**



M.Sc. THESIS

Muhammet Ali KADIOĞLU

Department of Industrial Engineering

Industrial Engineering Programme

JUNE 2022

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

**DISCOVERING MARKET INSIGHTS FROM ONLINE PRODUCT REVIEWS
THROUGH SENTIMENT ANALYSIS**

M.Sc. THESIS

**Muhammet Ali KADIOĞLU
(507191120)**

Department of Industrial Engineering

Industrial Engineering Programme

Thesis Advisor: Asst. Prof. Dr. Erkan Işıklı

JUNE 2022

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

**ÇEVİRİMİÇİ MÜŞTERİ YORUMLARI İLE DUYGU ANALİZİ VE PAZAR
HAKKINDA BİR İÇGÖRÜ ARACI**

YÜKSEK LİSANS TEZİ

**Muhammet Ali KADIOĞLU
(507191120)**

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

Tez Danışmanı: Dr. Öğr. Üyesi Erkan Işık

HAZİRAN 2022

Muhammet Ali Kadioğlu, a M.Sc. student of ITU Graduate School student ID 507191120, successfully defended the thesis/dissertation entitled “DISCOVERING MARKET INSIGHTS FROM ONLINE PRODUCT REVIEWS THROUGH SENTIMENT ANALYSIS”, which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Asst. Prof. Dr. Erkan IŞIKLI**
Istanbul Technical University

Jury Members : **Asst. Prof. Dr. Cafer Erhan BOZDAĞ**
Istanbul Technical University

Assoc. Prof. Dr. Nezir AYDIN
Yıldız Technical University

Date of Submission : 03 June 2022
Date of Defense : 27 June 2022





To my family,



FOREWORD

This thesis is written for my master's degree in Industrial Engineering. The subject of the study is related to Sentiment Analysis and Marketing Research. These are mutually complementary topics. The research aims to present a blend of technical and social sciences ranging from natural language processing to customer behaviors on digital platforms.

I would like to thank all the contributors to this study. Especially, I am really grateful to my thesis advisor Asst. Prof. Dr. Erkan IŞIKLI for providing guidance and feedback throughout the thesis.

I would particularly like to thank Advanced Analytics Department at Doğuş Teknoloji in Istanbul for all their support. Besides, special thanks to all the colleagues from Doğuş Teknoloji R&D Center for their support throughout my master's studies.

Finally, I am grateful to my family, and I owe my thanks to them for their endless love and support throughout my life.

June 2022

Muhammet Ali Kadioğlu
(Industrial Engineer)



TABLE OF CONTENTS

	<u>Page</u>
FOREWORD	ix
TABLE OF CONTENTS.....	xi
ABBREVIATIONS	xiii
SYMBOLS	xv
LIST OF TABLES	xvii
LIST OF FIGURES	xix
SUMMARY	xxi
1. INTRODUCTION.....	1
1.1 General Overview of E-commerce.....	2
1.2 Marketplace Trends.....	3
1.3 Customer Relations	4
1.4 Customer Satisfaction.....	5
1.5 Customer Loyalty	6
1.6 Customer Based Metrics	6
2. LITERATURE REVIEW.....	9
2.1 Text Analytics	9
2.1.1 Text Classification	9
2.1.2 Input representaiton.....	10
2.1.3 Deep learning	12
2.1.4 Transfer learning	14
2.1.5 Attention mechanism	14
2.1.6 Transformers architecture	15
2.1.6.1 Encoder	16
2.1.6.2 Decoder	17
2.1.6.3 Linear and softmax layer.....	18
2.2 Marketing Analytics	18
2.2.1 Market share.....	18
2.2.2 Share of requirements	19
2.2.3 Share of wallet	19
2.2.3.1 Individual wallet share	20
2.2.3.2 Aggregate wallet share	20
2.3 Related Studies	20
2.3.1 Working areas	21
2.3.2 Data sets	22
2.3.3 Data annotation methods.....	23
2.3.4 Classification algorithms.....	24
3. RESEARCH DESIGN AND METHODOLOGY	25
3.1 Research Objectives	25
3.2 Research Model.....	25
3.3 Sentiment Classifier – BERT	25
3.3.1 Tokenization.....	26
3.3.2 Modelling	28

3.3.2.1 Pre-Training tasks	28
3.3.2.2 Fine-Tuning tasks	29
3.3.3 Performance evaluation	29
3.4 Analyses	30
3.4.1 Sentiment analysis	30
3.4.1.1 Detail levels in sentiment analysis	31
3.4.1.2 Challenges of Turkish	31
3.4.2 Wallet allocation rule	32
4. EXPERIMENTS.....	35
4.1 Sampling and Data Collection.....	35
4.2 Determination of Product Categories	35
4.3 Data Annotation.....	36
4.4 Configuraitons	36
4.5 Data Pre-processing.....	37
4.6 Splitting the Dataset	39
4.7 Tokenization, Pre-Training and Fine Tuning	40
4.8 Making Predictions.....	41
4.9 Wallet Share	42
5. FUTURE WORK & CONCLUSION.....	47
5.1 Future Work.....	47
5.1.1 Model retraining.....	47
5.1.2 Aspect-based sentiment analysis	48
5.2 Conclusions	49
REFERENCES	51
APPENDICES	55

ABBREVIATIONS

MAE	: Mean Absolute Error
R²	: Coefficient of Determination
MAPE	: Mean Absolute Percentage Error
MSE	: Mean Squared Error
B2C	: Business to Consumer
B2B	: Business to Business
GDP	: Gross Domestic Product
CAGR	: Combined Annual Growth Rate
NPS	: Net Promoter Score
RNN	: Recurrent Neural Network
LSTM	: Long Short-Term Memory
SCR	: Share of Category Requirement
iSCR	: Individual Category Requirement Share
ROI	: Return of Investment
MDM	: Masked Language Model
NSP	: Next Sentence Prediction



SYMBOLS

TP	: True-Positive
FP	: False-Positive
FN	: False-Negative
TN	: True-Negative
$\bar{h}_t$	: Hidden State
$\bar{x}_t$	: Input Vector
t	: Time
P	: Precision
R	: Recall
j_0	: The Business Sales Volume
i_0	: The Customer Purchase Volume
j	: The seller serving in the market
i	: The customer purchasing in the market
J	: All sellers providing services or selling products in the market
I	: All of the customers who purchased services in the market
S_{ij}	: Total amount of sales made by seller j to customer i
V_{ij}	: Customer i 's Purchase Volume from Seller j
PB	: The relative order of preference a customer gives to a seller compared to other sellers used in the same category
TB	: Total number of sellers a customer uses in the category



LIST OF TABLES

	<u>Page</u>
Table 3.1 : Customer Satisfaction Levels and Share of Wallets (Keiningham et al, 2015)	33
Table 4.1 : Customer Ratings and Corresponding Classes.	36
Table 4.2 : The Weighted-Averaged F1 Scores.	41
Table 4.3 : Wallet Share Performance Metrics (Three-Class Approach).	44
Table 4.4 : Comparison of the Actual and Predicted Wallet Shares (Three-Class Approach with 500 Customers).	45
Table A.1 : Wallet Share Performance Metrics (2 Class – Type 1).	56
Table A.2 : Wallet Share Performance Metrics (2 Class – Type 2).	57
Table A.3 : Wallet Share Performance Metrics (5 Class).	58
Table B.1 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 1 & 100 Cus).	59
Table B.2 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 1 & 300 Cus).	60
Table B.3 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 1 & 500 Cus).	61
Table B.4 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 2 & 100 Cus).	62
Table B.5 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 2 & 300 Cus).	63
Table B.6 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 2 & 500 Cus).	64
Table B.7 : Comparison of Actual and Predicted Wallet Shares (3 Class & 100 Cus).	65
Table B.8 : Comparison of Actual and Predicted Wallet Shares (3 Class & 300 Cus).	66
Table B.9 : Comparison of Actual and Predicted Wallet Shares (5 Class & 100 Cus).	67
Table B.10 : Comparison of Actual and Predicted Wallet Shares (5 Class & 300 Cus).	68
Table B.11 : Comparison of Actual and Predicted Wallet Shares (5 Class & 500 Cus).	69



LIST OF FIGURES

	<u>Page</u>
Figure 2.1 : An outlook of the attention mechanism for a specific word (Vaswani et al, 2017).	15
Figure 2.2 : Transformer - Model Architecture (Vaswani et al, 2017).	16
Figure 3.1 : BERT's Solution Architecture for Text Classification (Devlin et al, 2019)	26





DISCOVERING MARKET INSIGHTS FROM ONLINE PRODUCT REVIEWS THROUGH SENTIMENT ANALYSIS

SUMMARY

Online shopping is increasingly preferred, products and services are compared, and customer reviews are shared on numerous digital platforms thanks to the successive iterations of the web and the widespread of the Internet. This type of customer feedback is of great value for marketing activities. As the content created on digital platforms is easily accessible and the sentiment of text data can be analyzed using various learning algorithms, marketers can conveniently produce market intelligence about product and service perceptions and customer behaviors.

Due to the continuously expanding volume of customer evaluations, approaches based on manual analysis are insufficient to provide a broad perspective of consumer opinions for a product. The need to automate the process has led to the development of various analysis tools. As a result, it has become necessary to build data analysis systems that can automatically evaluate customer reviews and summarize their behavior and opinions.

Using natural language processing tools and deep learning algorithms to extract insights from digital platforms, sentiment analysis can identify the emotional value of communications for businesses and help marketers understand their customers better. More effective marketing strategies can be designed using unstructured data. Accordingly, analyzing unstructured data to identify aspects, sentiments, emotions, keywords, and purposes can help to improve the market research of a business. Furthermore, marketers may also use data analysis to find customer pain points, determine their bottlenecks, and track their competitors.

This study essentially presents a systematic review of text and market analytics. It also delivers detailed findings and theoretical backgrounds of Turkish NLP applications, transformers, and market research methods. Firstly, the rapid growth of digital platforms and e-commerce under technological improvements and pandemic conditions is acknowledged. During the lockdown, how e-commerce, which had already experienced a significant growth trend, accelerated and grew is explained. Secondly, how to determine the metrics that measure consumer habits and behaviors in digital marketplaces are discussed. Macro-level indicators related to e-commerce volume are given for both Turkey and the world. A framework for which factors have affected e-commerce recently and which trends it can follow in the future is presented. Thirdly, research on how customer satisfaction and loyalty can be measured considering a customer's experience on digital platforms is conducted. Issues that affect and shape customer relations in digital channels are defined, and which metrics to follow for their measurement is discussed. Finally, the first chapter investigates how customer satisfaction and loyalty can improve customer experience in e-commerce platforms and their impact on profitability.

The literature related to text analytics and market research is thoroughly reviewed. In the text analytics section, the classification of texts is primarily discussed. Sentiment analysis involves the task of assigning text data to a sentiment class. The methods used in text classification are mentioned, and the ways to evaluate classification performance are explained. Subsequently, how to represent text data that is to be introduced as an input to an NLP model is explained. The word embeddings used to describe text data are mentioned, and their positive and negative aspects are explained. An example discussing how two words with the same meaning can make a difference in word embeddings is provided. Natural language processing problems are sought to be answered by deep learning. The fundamentals of deep learning are explained in detail, and how it can work with text data is presented. The difficulties of working with text data are defined, and the methods developed for deep learning algorithms to overcome these difficulties are described. Using the knowledge and experience gained from similar natural language processing problems instead of learning from scratch is examined under the heading of transferred learning. Finally, the theoretical background of attention mechanisms, which are frequently used in the solution of NLP problems, is explained. The solution proposed by the transformers is examined.

Sentiment analysis helps businesses monitor their customers' perceptions of their products and provides market intelligence about the ecosystem. Online product reviews may be a significant data source for gaining market insights and analyzing the competitors. It also allows for learning what customers in the target segment think about the competitors that are operating in the same market. These deep insights help fill gaps in products and services and help sellers stay ahead of the competition. In addition, understanding customers' experience on digital platforms provides information about customers' satisfaction levels and enables to monitor their interactions with the digital channels. In addition, sentiment analysis helps one evaluate how a customer group perceives a product or service.

Diversifying digital channels used by online platforms shape the marketing analytics processes. Market strategies are reshaped with insights gained from the market using analytical tools. These tools can be used effectively for businesses to achieve their operational and financial goals with the help of marketing data. The first metric examined in marketing analytics within this context is the market share, which is one of the most critical metrics used by businesses to determine their current situation and the market targets to be followed. In addition, the need share and the wallet share metrics, which focus on customers and show the share of businesses in customer expenditures, are explained. Two different approaches to the wallet share, where customers are examined individually and evaluated cumulatively, are also discussed.

The literature on Turkish sentiment analysis has generally employed traditional machine learning methods; however, models using deep learning and transformative architecture have recently become more popular. Sentiment analysis has been studied in various fields, using movie reviews, hotel reviews, social media shares, and online marketplaces reviews. The size of data sets, the way data annotation is performed (manually vs. automatically), and the data preprocessing methods are the main differentiators of these studies. Consequently, the end-to-end sentiment classification task is examined step by step.

Firstly, how tokens represent text data for the BERT model, used as a sentiment classifier, and how the word embedding method used means the semantic relationship in the text is described. BERT's solution proposals, especially for text classification

tasks, are shown in detail. Then, how the pre-training and fine-tuning tasks perform the learning process using attention mechanisms in the modeling phase, and performance monitoring metrics are explained.

The level of detail of the sentiment analysis is given, and difficulties encountered in the natural language processing problems in an agglutinative language, Turkish, are explained. The wallet allocation rule and how the sellers can use it to determine a better market strategy is explained. Furthermore, how the wallet share can affect the financials and how the enterprises can turn this information into an advantage knowing this effect is discussed.

Experiences are gained using customer reviews on an e-commerce marketplace for online sellers within the scope of this study. Data analysis, how product categories are determined, data annotation, adjustments to the data set, data preprocessing, identifying testing and training data sets, tokenization of text data, modeling, generating predictions, and performance monitoring processes are described. The model's outputs, estimation results, are introduced as inputs to calculate the wallet share. Finally, the wallet shares that the customers have allocated to the sellers are counted.

Different multi-class classification model alternatives (5-class, 3-class, and 2-class) are evaluated, and the 3-class classification is eventually selected. The 3-class classification weighted-averaged f1 scores are 0.8631 for the category of case, 0.8626 for men's perfume, and 0.9016 for mobile phones. Since the wallet allocation rule calculates the customer-based wallet share, the reviews are randomly assigned to the customers. The models are developed using different customer sizes (100, 300, and 500) to see the effects on estimating the wallet share. The model developed with the biggest customer size gives the lowest absolute error. The assumption can be summarized that the wallet shares calculated with the online reviews - collected from three categories, grouped into three sentiment classes, and received from 500 customers. Needs for improvement are specified, and a brief guideline is identified for topics that can be considered in future work. The learning model's automatic retraining structure is proposed, and aspect-based sentiment analysis is described.



ÇEVİRİMİÇİ MÜŞTERİ YORUMLARI İLE DUYGU ANALİZİ VE PAZAR PAYI İÇİN BİR İÇGÖRÜ ARACI

ÖZET

Çevrimiçi alışveriş giderek daha fazla kullanılmakta, insanlar satın alacakları ürün ve hizmetleri karşılaştırıp, deneyimleri hakkında yorum yapmaktadır. Her geçen gün hayatımızda daha çok yer edinen dijital platformlar aracılığıyla yapılan değerlendirmeler paylaşılmaktadır. Bu tür müşteri geri bildirimleri, pazarlama zekası uygulamaları için değerli bir kaynaktır. Dijital platformlarda oluşturulan içeriklere kolayca erişilebilmekte ve çevrimiçi ürün yorumları, öğrenme algoritmaları kullanılarak analiz edilebilmektedir. İşletmeler sundukları ürün ve hizmetler için değer yaratabilecek önemli bir imkana sahiptirler.

Çevrimiçi platformlarda sürekli artan müşteri değerlendirmeleri nedeniyle manuel yaklaşımlar, analiz için yetersiz kalabilir. Süreci otomatikleştirme ihtiyacı, analiz araçlarının gelişimini de beraberinde getirmiştir. Böylece, müşteri yorumlarını otomatik olarak değerlendirebilen veri analiz sistemleri oluşturmak gerekli hale gelmiştir. Dijital platformlardan içgörü elde etmek için doğal dil işleme araçlarını ve derin öğrenme algoritmalarını kullanan duygu analizi, işletmelerin müşteri ile arasındaki iletişimin duygusal değerini belirleyebilir ve müşteriler hakkında daha fazla bilgi sahibi olmalarına yardımcı olabilir. Üretilen toplam verinin %80'ini temsil eden yapılandırılmamış veri kullanılarak etkili pazarlama stratejileri tasarlanabilir. Duyguları, anahtar kelimeleri ve hedef alanları belirlemek için yapılandırılmamış veriyi analiz etmek, işletmelerin pazar araştırmaları için etkili bir araç olarak konumlandırılabilir. Öte yandan pazarlamacılar, müşterilerinin sorun yaşadıkları alanları bulmak, darboğazları belirlemek ve pazardaki rakiplerinden farklılaşabilmek için duygu analizini kullanabilir. Bu çalışmada, duygu ve pazar analizinin sistematik bir değerlendirmesi sunulmuştur. Geliştirilen hibrit çözümün bulguları paylaşılmış ve bir Türkçe doğal dil işleme uygulaması gerçekleştirilmiştir.

Teknolojik gelişmeler ve COVID-19 pandemisi şartları ile dijital platformların ve e-ticaretin hızla büyümesinden bahsedilmiştir. Halihazırda iyi bir büyüme trendi yaklayan e-ticaretin, pandemi nedeniyle herkesin evlere kapanması ile nasıl ivmelenerek büyüdüğü üzerinde durulmuştur. Giderek hayatımızda daha büyük yer edinen dijital pazaryerlerindeki müşteri deneyimi ve davranışları tartışılmıştır. E-ticaret hacmi ile ilgili Türkiye'ye ve Dünya'ya ait makro göstergeler incelenmiştir. Son dönemde e-ticaretin hangi etmenlerden etkilendiği ve gelecekte hangi trendleri takip edebileceği hakkında bir çerçeve oluşturulmuştur. Dijital platformlarda sunulan deneyim göz önünde bulundurularak, müşteri memnuniyeti ve sadakatının nasıl ölçülebileceği ile ilgili araştırmalar üzerinde durulmuştur. Dijital kanallarda işletmeler ve müşteri arasındaki ilişkiyi etkileyen ve şekillendiren hususlar tanımlanmış ve ölçümü için hangi metriklerin takip edilebileceği tartışılmıştır. Müşteri memnuniyeti ve sadakatının e-ticaret platformlarındaki deneyimi iyileştirmek için nasıl kullanılabileceği ve karlılık üzerindeki etkisi tartışılmıştır.

Literatür taraması, metin ve pazar analitiği olmak üzere iki ana başlık altında yapılmıştır. Metin analitiği bölümünde, ilk olarak metin sınıflandırılması incelenmiştir. Metin sınıflandırması yaparken kullanılan yöntemlerden bahsedilmiş ve sınıflandırma performansının nasıl değerlendirileceği açıklanmıştır. Metnin bir duygu sınıfına atanması görevini içeren duygu analizi için metin verisinin öğrenme modeline girdi oluşturacak şekilde nasıl hazırlandığı anlatılmıştır. Temsil için kullanılan kelime gömülmelerinden bahsedilmiş ve kelime gömülmelerinde aynı anlama gelen iki kelimenin nasıl temsil edileceği tartışılmıştır. Doğal dil işleme problemlerinin çözümünde kullanılan derin öğrenme ayrıntılı olarak açıklanmış ve metin verisi ile nasıl çalışabileceği anlatılmıştır. Metin verisi ile çalışmanın getirdiği zorluklar tanımlanmış ve derin öğrenme algoritmalarının bu zorluklar ile mücadele etmek için ne tür yöntemler geliştirdiği açıklanmıştır. İlk defa öğrenme yerine benzer doğal dil işleme problemlerinden edinilen bilgi ve birikimin nasıl kullanılabileceği aktarımlı öğrenim başlığı altında incelenmiştir. Son zamanlarda NLP problemlerinde sıkça kullanılan dikkat mekanizmalarının teorik arka planı anlatılmış ve dönüştürücü mimarisi ile sunduğu çözüm önerisi incelenmiştir.

Müşterilerin ürün ve hizmetlere ilişkin ne düşündüklerini göstermesinin yanı sıra, duygu analizinin pazar hakkında nasıl bilgi sağlayabileceği tartışılmıştır. Çevrimiçi ürün yorumları kullanılarak bir pazarlama aracının gelişimi gösterilmiştir. Bu araç, müşterilerin pazarda faaliyet gösteren rakipler hakkında içgörü kazanmasına imkan sağlayacak şekilde tasarlanmıştır. Ayrıca analiz sonuçları ürün ve hizmetlerdeki eksikleri tamamlamak için kullanılabilir ve satıcıların rekabette öne geçmelerine yardımcı olabilir. Müşterilerin dijital platformlarda yaşadıkları deneyimin öğrenilmesine imkan sağlar, müşterilerin memnuniyet düzeyleri hakkında bilgi verir ve dijital kanallar ile etkileşimin nasıl gerçekleştiğinin incelenebilmesine imkan sağlar. Bunların yanında duygu analizi, bir ürün veya hizmetin hedef ve/veya potansiyel müşteri kitlesi tarafından nasıl algılandığının değerlendirmesine yardımcı olur.

Çevrimiçi pazaryerlerinin müşterileri ile etkileşime geçtikleri dijital kanalların çeşitlenmesi, pazar analitiği süreçlerini şekillendirmektedir. Pazar analitiği, işletmelerdeki karar vericilerin müşteri odağında pazar stratejilerini şekillendirebilmeleri için bir araç olarak kullanılabilir. İşletmelerin pazarlama verisi ile operasyonel ve finansal hedeflerine ulaşmaları için etkili bir araçtır. Bu kapsamda pazar analitiği ile ilgili ilk incelenen metrik pazar payı olmuştur. İşletmelerin mevcut durumlarını ve takip edecekleri pazar hedeflerini belirlemeleri için kullandıkları en önemli metriklerden biri olan pazar payı incelenmiştir. Ayrıca odağına müşterileri alan ve müşteri harcamalarından işletmelerin aldıkları payı gösteren ihtiyaç payı ve cüzdan payı metrikleri açıklanmıştır. Bireysel ve toplu olmak üzere özelleşen cüzdan payı incelenmiştir.

Türkçe duygu analizi ile ilgili çalışmalar göz önüne alındığında, genelde geleneksel makine öğrenmesi metodları ile gerçekleştirildiği, son zamanlarda derin öğrenme ve dönüştürücü mimarisini kullanan modellerin de popülerleştiği görülmüştür. Film ve otel yorumları, sosyal medyadaki gönderiler, online platformlarda ürün ve hizmetlere yapılan yorumlar başta olmak üzere birçok alanda duygu analizi yapılmaktadır. Çalışılan alana göre çeşitli hacimlerdeki veri ile analiz çalışmaları gerçekleştirilmektedir. Hem manuel hem de otomatik olarak veri işaretleme süreci tamamlanmaktadır. Metin verisinin bir öğrenme modeline girdi oluşturacak şekilde nasıl hazırlandığına ait örnek çalışmalar da incelenmiştir. Duygu sınıflandırması görevi için kullanılan modeller ayrıntılı olarak aktarılmıştır.

Duygu sınıflandırıcı olarak kullanılan BERT modelinde, metin verisinin işlenebilir hale gelmesi için jetonlar ile nasıl temsil edildiği anlatılmıştır. Kullanılan kelime gömülmesi metodunun kelimeler arasındaki anlamsal ilişkiyi nasıl temsil ettiği tartışılmıştır. BERT'in metin sınıflandırma görevleri için sunduğu çözüm önerisi ayrıntılı olarak gösterilmiştir. Modelleme aşamasında ön eğitim ve ince ayar görevlerinin dikkat mekanizmaları ile öğrenme sürecini nasıl gerçekleştirdiği anlatılmıştır. Modelin değerlendirilmesi ve performans metriklerinin nasıl izlenmesi gerektiği hakkında bilgilendirme yapılmıştır.

Bu çalışma kapsamında gerçekleştirilen analiz yöntemleri, duygu ve pazar analizi, açıklanmıştır. Duygu analizinin ayrıntı seviyeleri aktarılmış ve sondan eklemeli bir dil olan Türkçe dilindeki doğal dil işleme problemlerinde karşılaşılan zorluklardan bahsedilmiştir. Cüzdan Payı Kuralı'nın ne olduğu, nasıl kullanıldığı aktarılmıştır. Cüzdan payının işletmelerin finansalları üzerinde nasıl bir etki oluşturabileceği ve öğrenilen bu etkiyi nasıl avantaja dönüştürebilecekleri tartışılmıştır. Tasarlanan hibrit çözüm, bir e-ticaret pazaryerindeki çevrimiçi müşteri yorumları kullanılarak uygulanmıştır. Veri analizi, ürün kategorilerinin belirlenmesi, veri işaretleme, veri setinin işlenebilmesi için yapılan ayarlamalar, veri ön işleme, veri setinin test ve eğitim olarak ayrılması, metin verisinin kelime gömülmeleri ile temsili, modelleme ve tahminlerin üretilmesi adım adım anlatılmıştır. Öğrenme modelinin ürettiği tahminler, cüzdan payı kuralını hesaplayan model için girdi oluşturmuştur. Müşterilerin satıcılara ayırdıkları cüzdan payları tahmin edilerek, satıcıların pazarlama aktivitelerinde kullanabilecekleri bir araç geliştirilmiştir.

Metin sınıflandırmasında 2-sınıflı, 3-sınıflı ve 5-sınıflı modeller değerlendirilmiş ve 3-sınıflı sınıflandırma ile ilerlemeye karar verilmiştir. Kılıf kategorisi için tahmin başarısı 0.8631, erkek parfümü için 0.8626 ve cep telefonu için 0.9016 olarak bulunmuştur. Cüzdan payı kuralı, müşteri bazlı cüzdan payını hesapladığı için veri setindeki müşteri yorumları oluşturulan müşterilere rastgele atanmıştır. Atanan müşteri sayısının cüzdan payı tahminindeki etkisinin araştırılması için 100, 300 ve 500 adet müşteri için rastgele atama işlemi gerçekleştirilmiş ve sonuçlar değerlendirilmiştir. En yüksek müşteri sayısının, en düşük hata ile tahmin yapılabilmesine olanak sağladığı görülmüştür. Üç kategori için toplanan müşteri yorumlarının her kategoride 500 müşteriden alındığı varsayılarak yorumlar müşterilere rastgele dağıtılmış ve satıcılara ayırdıkları cüzdan payları hesaplanmıştır. Bu çalışmanın geliştirilmesi ve iyileştirilmesi için gelecek dönemde göz önünde bulundurulabilecek konular hakkında bir ön araştırma da yapılmıştır. Duygu analizinde kullanılan öğrenme modelinin otomatik ve düzenli olarak eğitilebileceği bir yapı önerilmektedir. Ayrıca çevrimiçi müşteri yorumlarının hangi deneyim alanını hedeflediğinin tespiti ve hedefe ait duygu sınıfı ataması ile daha ayrıntılı bir analizin yapılabilmesi üzerinde durulmuştur.



1. INTRODUCTION

E-commerce and the digital economy are becoming increasingly significant with today's technological developments and the effects of COVID-19 pandemic. It has become common for end-users to share their experiences about the services and products through various platforms. This transfer of knowledge has an important place in the purchasing decisions of new users. Businesses that are aware of the importance of monitoring the information shared by their customers put an effort to take positive customer reviews as a goal of their marketing operations. Apart from new opportunities, this process also brings about new requirements and new challenges.

Consumers have been spending more time at home due to the restrictions imposed by COVID-19 pandemic. The number of digital platform users and usage times have thus increased (Göl et al, 2022). Moreover, digital platforms transform the way companies and consumers communicate and interact around the world. The acceleration of digital platforms and e-commerce has brought the need to analyze their data. Data-driven insights have provided a better understanding of local and international e-commerce trends and an easy adaptation to the transformation. COVID-19 pandemic engendered a dramatic slowdown in economic activity worldwide. Demand for digital platforms in e-commerce eventually increased. The trend for e-commerce observed during the pandemic is expected to continue during the recovery period (United Nations, 2021).

A customer journey starts at the moment of first touch in the digital marketplace. Providers use various channels to understand and analyze end-to-end customer experience on their platforms. Thoughts on the product or service are written as texts. Sellers can quickly learn from the customer experiences and investigate the sentiments using the product reviews left on their platforms. If there are one or two products but not many customers, the analysis could be completed manually. However, nowadays, digital marketplaces offer more products than their number of customers. Consequently, there is a need to automate the process using natural language processing models to extract ideas from the reviews, make sense of the expressions within, and gain relevant insights to achieve strategic goals. Analyzing the sentiments

of product reviews at digital marketplaces is vital to improve the provided service and the quality of products, ensure customer loyalty, understand the market, and increase the brand value.

1.1 General Overview of E-commerce

In 2020, China had the greatest sales volume in the e-commerce market (\$2.29 trillion). The United States came second with total sales of \$794.50 billion, and the UK was third with total sales of \$180.39 billion. Other strong players in the market were Japan (\$141.26 billion), South Korea (\$110.60 billion), and Germany (\$96.86 billion) (Göl et al, 2022). The e-commerce volume, which had been increasing steadily, has been continuing to grow exponentially since the emergence of COVID-19 pandemic. Business-to-Customer (B2C) e-commerce turnover and the share of GDP made up by e-commerce have been increasing in European countries. (Lone et al, 2021)

Economies that leverage the potential of e-commerce are better positioned, while those that do not are lagged behind. Most economies are positively affected by the global growth in e-commerce. Developing countries should pay special attention to this potential. The e-commerce expenditures realized in Turkey between 2016-2020 increased by 3.2 times and reached 226.2 billion TL according to the values in 2020 (Göl et al, 2022). This growth shows a combined annual growth rate (CAGR) of 43% over the four years. Besides, in 2020, the Turkish B2C e-commerce market grew 29.6% in terms of local currency, reaching a volume of 101.5 billion TL (approximately US\$ 11.9 billion). This growth rate is higher than the rates achieved in the last three years. Therefore, e-commerce provided an excellent opportunity to overgrow during the pandemic.

On the other hand, such rapid rises come from a low e-commerce baseline. The fact that 60% of Turkish citizens have never shopped online indicates an enormous potential for growth in the Turkish e-commerce sector (JPMorgan Chase & Co., 2021). During the pandemic, online grocery shopping increased fivefold. Turkey has a significant advantage compared to more developed e-commerce economies in the Western World regarding application-based innovations due to its young population. Furthermore, the authorities have begun to regulate and actively monitor social commerce advertising and marketing.

Especially in 2020, with the positive impact of the pandemic in the e-commerce sector, a significant increase was observed in the expenditures both in Turkey and in the rest of the world. The numbers provide a positive view of the potential and trend of the Turkish e-commerce market. The transformation in consumer behavior during the pandemic period is expected to be permanent. Hence, e-commerce will be one of the most significant strategic goals of businesses. In the Turkish market, local players continued shaping e-commerce. The most visited e-commerce platforms are local marketplaces such as Trendyol, Hepsiburada, and n11.com. Global players such as Amazon also operate in Turkey.

In addition to the growth in e-commerce sales, the variety of products and services it covers is also expanding. Business-to-business (B2B) e-commerce generally includes products from the manufacturing industry and wholesale trade. In retail e-commerce, clothing comes first among the most purchased products and services. The clothing category is followed by the shoes and consumer electronics categories (Deliçay, 2021). The growth trend in retail e-commerce, which highlights consumer products, draws attention to multi-category e-marketplaces that have almost all these categories in their portfolios (Sarıçiçek et al, 2022). As a result, e-marketplaces are gaining prominence as the rising stars of retail e-commerce, both in the world and in Turkey.

1.2 Marketplace Trends

Trade barriers have decreased, and both established and emerging trade networks have redefined the geographic boundaries. Thus, location is no longer seen as an advantage for businesses. Globalizing markets have increased the demand for logistics management and cargo operations. Moreover, the quality of product attributes in the market has increased significantly, which no longer provides a competitive advantage for most companies. A qualified product needs more to compete in a market with high standards. As the quality differences of products decrease, businesses try to gain a competitive advantage through experience-oriented relationships.

Additionally, the brand loyalty a product provides by differentiating in the market does not ensure absolute success. It is a relative success since the reaction of customers to products and services is becoming increasingly personalized (Kumar and Reinartz, 2018). To obtain customers' opinions about products or services and create value from

them by considering the impact of personal experience on purchasing behavior is thus vital for businesses.

1.3 Customer Relations

The growing e-commerce usage has brought along the need for digital transformation. This transformation does not solely and exclusively mean the digitization of business processes. Several innovations are also required in the business processes, and there is a need to revise the view of customer relations. A customer's value is not defined only by the revenue stream from purchasing goods and services. Following their purchase, customers can become substantial brand representatives for businesses. Online platforms, which have a great place in our lives, are actively used to express personal opinions about a product or a brand. Related examples would be the acquisition of new customers based on the recommendations of existing customers or negative product reviews that are likely to prevent potential customers from purchasing a product.

Defining customer loyalty and seeking ways to improve it has an essential place in the management of customer relations in e-commerce applications. Kumar et al. (2010) examined a customer's loyalty to a business with respect to four metrics. The first metric is the customer lifetime value which represents the purchasing behavior of customers. The second metric is the customer referral value, which addresses the behavior of potential leads. The third one is the customer influencer value. It has been shown that a customer provides value to the business that influences other customers. Finally, the fourth one is the customer knowledge value. It aims to provide value added to the business based on the feedback from the customer. These four metrics can help digital platforms evaluate all aspects of the experience they offer to their customers.

Businesses can follow their performance management using various metrics while preparing their customer relationship strategies. Metrics can be renewed depending on the product and service offered, and these targets can be customized considering the strengths of the businesses' marketing departments. In addition, they can decide their position in the market while developing strategies to increase customer loyalty and draw up action plans within this context for the functions they think require development.

1.4 Customer Satisfaction

Understanding and meeting customer needs have become the primary dimension in which companies create a competitive advantage by analyzing marketing activities at the individual level (Kumar and Reinartz, 2018). Every touchpoint between the product/service and the customer provides an attainable resource to generate a close relationship. Moreover, customers play an active role in marketing communications by providing feedback to the business, reviewing the product for the use of other customers, and so on. Thus, providers have been renewing their marketing strategies from product-based to customer-centric.

The direct effect of marketing investments on the company's profitability increases the pressure on businesses to renew their marketing strategies (Kumar and Reinartz, 2018). At the same time, explaining why marketing expenditures are necessary is significant in terms of accountability to stakeholders. However, the expansion of communication channels and the decrease in efficiency and effectiveness overburden this account. Therefore, it becomes a crucial goal for marketing professionals to define, track and improve the metrics they can be related to the profitability of their businesses.

Research on how to improve customer experience is not in its infancy. However, technological developments have helped to diversify and expand the data sources. Many channels have been digitized, and data acquisition has become more accessible. The role of improving customer experience in satisfaction has increased the interest in studies on this subject, and the developing literature has brought complexity. Satisfaction reflects the thoughts covering the entire consumption experience and describes pleasure, which is an essential emotional response for customer loyalty (Piris and Gay, 2021). The relationship between customer satisfaction and financials has increased the providers' attention to the studies on this subject.

Studies on customer satisfaction aim to determine the non-linear relationship between various exactly nonproper features and satisfaction or to understand the level of satisfaction from a general perspective through indirect measures such as the net promoter score (NPS). While companies use NPS to assess overall satisfaction, its capacity to predict growth has been questioned (Piris and Gay, 2021; Keiningham et al, 2007).

1.5 Customer Loyalty

Customer loyalty is a crucial strategic goal that must be achieved for all product and service providers. Traditional studies on customer loyalty revolve around customer retention activities. Customer satisfaction has been the most commonly used metric to measure and manage customer loyalty. Businesses need to focus on the share of their customers in their wallets by interpreting customer satisfaction levels for the products or services they offer. Coolil et al. (2007) reported a positive non-linear relationship between customer satisfaction and wallet share. The demographic characteristics of customers and the situational factors that influence customers were also examined as features affecting the relationship between customer satisfaction and wallet share. The effect of customer satisfaction on the wallet share decreases as the customer income and the duration of the relationship with the company increase.

Previous studies reported that loyal customers allocated a higher share of category spending to the business and were more likely to recommend to other potential customers (Keiningham et al, 2007). Thus, providers aim to shape their marketing strategies by measuring loyalty to identify their customers. In addition, businesses use customer feedback to measure their customers' satisfaction and loyalty so that they can monitor their performance in achieving their strategic goals (Morgan and Rego, 2006).

1.6 Customer Based Metrics

Companies can prioritize customers they predict would generate more profit with customer-based value metrics to allocate their resources efficiently. On the other hand, it should be ensured that the information these metrics provide to the business regarding customers' purchasing behavior is correct. The metrics that are highly correlated with customers' purchases should be closely monitored. The size of the wallet shows the total amount of purchases made by a customer from companies operating in the market for a specific product category. Businesses compete for the largest share of the customers' wallets. To get ahead in this competition, they must know where they are and where they want to be.

Marketing research could be conducted through a survey or other tools. For instance, existing customers can be asked how much they spend each month in Category A. However, obtaining the wallet size information of customers is challenging. Besides,

information in customer segments would be general and cannot provide enough detail. Businesses, thus, may not be able to learn the size of their customers' wallets directly at the individual level. However, they can specify an alternative metric that is highly correlated with this value. Hence, it should not be forgotten that the possibilities and capabilities of the enterprise will affect the accuracy and quality of the information it will obtain.

A metric determined by asking customers how likely they are to recommend the product or service to others, Net Promoter Score (NPS), is argued to be the best customer loyalty metric (Reichheld, 2003). In another study, Piris and Gay (2021) identified the issues in customer feedback through Latent Dirichlet Allocation. The authors used an application that automated the processing of customer reviews and Net Promoter Score to represent the level of satisfaction (Piris and Gay, 2021).

On the contrary, Morgan and Rego (2006) argued that the net promoter score, which indicates customer intent, and the average number of recommendations, which track customer behavior, had no effect on business performance. Furthermore, they emphasized that focusing solely on customers' referrals and behaviors on feedback systems would mislead businesses (Morgan and Rego, 2006).

Keiningham et al. (2007) stated that there was no consensus on the best way to measure customer loyalty. The authors also asserted that no peer-reviewed scientific studies examined the relationship between recommendation intention and customer behavior (Keiningham et al, 2007).

Keiningham et al. (2015) revealed that there was no connection between customer satisfaction and market share/wallet share. As shown Table 3.1, the author argued that customer satisfaction could explain only 1 percent of customers' category spend.

In addition, they stated that commonly used measures of customer loyalty, such as NPS, failed in this regard. This study indicates that performance management systems upon which businesses have built marketing strategies should be reviewed, revised, or reconstructed.

Section 2 explains the theoretical background of the methods used in this study in detail. Next, the proposed solution, the research design, and the methodology are defined. Finally, an application of the proposed methodology is delivered using online customer reviews of a digital platform serving in the field of e-commerce. A general

evaluation follows, accompanied by a brief discussion on possible avenues for future research.



2. LITERATURE REVIEW

2.1 Text Analytics

More than 80% of the information currently available on the internet is estimated to be unstructured (Giudice et al, 2019). Text analytics offers a multi-disciplinary approach, using machine learning, statistics, and linguistics methodologies to extract insights from vast amounts of unstructured text data with no specific format. As a result, it allows businesses, organizations, and researchers to make critical choices using the voluminous data available.

Text analytics can work with text data obtained from various channels, including e-mails, online posts, surveys, and phone calls. Sentiment analysis, one of the most commonly used techniques in text analytics, can help determine how people feel about a product or understand the most critical issues in their reviews. On the other hand, it can allow decision-makers to detect a sudden increase in negative reviews. It provides invaluable information to improve the quality of the product/brand/service, such as resolving a defect causing customer dissatisfaction.

2.1.1 Text Classification

Classical methods used in binary classification problems include algorithms such as Logistic Regression and Support Vector Machines. In addition, multi-class classification methods such as Random Forest and Naive Bayes are used when the classification problem involves more than two classes. However, multiple binary classifiers can be used for multi-class classification.

In multi-class classification, expressions are assigned to a class. With multi-label classification, any expression can be assigned to more than one class. Model success can be evaluated by calculating the F1-score for each label and averaging these scores over the labels. When averaging, the labels are assumed to be equally important. The weights of the labels can be determined according to the problem, and the performance is measured using weighted averages (Géron, 2019).

Confusion matrix is used to evaluate the accuracy of a model built for a classification problem. It allows one to measure how accurately the model performs the classification task by comparing the predicted values with the actual ones. The actual and predicted values in the confusion matrix create performance criteria such as accuracy, recall, precision, and F1-score and they are calculated by the following formulas.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2.2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2.3)$$

$$\text{F1 - score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2.4)$$

True-Positive (TP), False-Positive (FP), False-Negative (FN), and True-Negative (TN) are values that express the relationship between the actual observations and predictions. Positive and negative conditions represent the number of real negative and real positive cases in the data. Besides, test results that correctly or wrongly indicate the presence of a condition are used to define the performance of the classification.

2.1.2 Input representaiton

In natural language processing applications, text data can be separated into words, syllables, characters, or n -grams (Brown et al, 1992) and converted into vectors. Texts are tokenized and can be associated with numeric tensors via vectors. Thus, text data are made available as input for the learning model. In neural networks, data representation is made with arrays and array sets of various sizes, called tensors, which act as carriers for data and generally contain numerical data. Deep neural networks contain many interconnected tensor operations.

In natural language processing applications, the first step in the preprocessing stage may be to convert the text data to tensors and obtain the word vectors of the text. Widely used normalization processes reduce the range of values and ensure homogeneity between existing features. In addition, transforming all features so that they have a mean of 0 and a standard deviation of 1 and they take values from the interval $[0,1]$ are frequently used to help learning algorithms work more efficiently.

One of the most common examples of associating a vector with tokens is word embeddings. Word vectors can be obtained by starting with random vectors and learning the network's weights. Another method uses pre-trained word embeddings for a similar natural language processing task. Converting a text into a word embedding, the semantic structure of the text is modeled numerically. In this way, it becomes possible to perform mathematical operations using word embeddings that represent the text.

The relationships between words can be shown in the representations. The semantic similarity of words is indicated as the distance between vectors representing the words. While creating word embeddings, the relationship between the meanings of the words can be expressed by the geometric distance between the word vectors. Also, the direction of vectors have significance in the embedding space. For instance, synonymous words that can be used interchangeably in sentences must be represented in word embeddings. Besides, pre-trained word embeddings can be used as an alternative. If the ready-to-use embedding vectors are available, they could provide the vectors which represent the domain knowledge and the general structure of language (Chollet, 2017).

To better understand word embeddings, we can examine the following two sentences as examples.

*Sipariş ettiğim ürün çok **hızlı** geldi. (The product I ordered arrived very **quickly**.)*

*Sipariş ettiğim ürün çok **çabuk** geldi. (The product I ordered arrived very **rapidly**.)*

The two synonyms that are written with bold letters must be represented in similar places in the vector space. That is, there should be a very short distance between these word vectors.

Even though this is not true, all words are represented differently from each other using one-hot encoded vectors. In this example, the words “hızlı” (quickly) and “çabuk”

(rapidly) are synonyms and thus interchangeable. However, "hızlı" (quickly) has no similar semantic relation to "ürün" (product). Therefore, such semantic relations must be shown in the vector space so that the model can learn the meaning in the given text. In other words, the difference between words that have synonyms or similar definitions and words that have different meanings should be representable in the vector space.

It has been discussed whether words are best represented with count-based methods or prediction-based methods. In this regard, Baroni et al. (2014) compared these two methods in lexical semantics tasks and claimed that predictive models outperformed. On the other hand, Pennington et al. (2014) introduced GloVe and argued that context-counting and context-predicting methods were fundamentally similar. Besides, they emphasized that context-counting methods better captured global statistics, which expressed the meaning of the sequence of words in the text. Nowadays, predictive models are frequently used in natural language processing tasks. Methods such as word2vec (Mikolov et al, 2013), which makes predictive word representations via neural networks, have been preferred more.

2.1.3 Deep learning

Machine learning algorithms generally perform learning operations using one layer. Deep learning gains more depth with successive layers and tries to find the best solution by adopting a layered architecture. Since the first input is selected from random values, the neural network's prediction will be far from the actual value. The network uses this loss as feedback to update the weights. Each time the flow is examined, the weights are updated. Thus, each iteration produces more accurate predictions (Chollet, 2017).

Deep learning methods depend on the type of the problem, the structure of the data, and so on. For example, Recursive Neural Networks have generally been developed in text analytics or with sequential data such as time series and biological series. The point that makes RNN stand out when working on texts is its success in understanding the capability of sequence structures. While taking each element of the arrays as an input, it also keeps the state information and uses it to produce the next output. RNNs can simultaneously consider each input with others. Instead of making the whole data available at once, sequentially introducing it as an input to the model allows the state information to be retained. RNN tries to capture the relations between the expressions

in the text using the hidden layers created from previous inputs. The class assignment output is produced only through the last input in the text classification models. The hidden state ($\bar{h}_t$) at (t) is expressed by a function of the hidden vector ($\bar{h}_{t-1}$) at ($t - 1$) and the input vector ($\bar{x}_t$) at (t) (Aggarwal, 2018):

$$\bar{h}_t = f(\bar{h}_{t-1}, \bar{x}_t) \quad (2.5)$$

While reading a text, we move forward keeping in mind what we have read before. How deep learning models extract the meaning from the text is already defined. However, keeping all the information in the text would get more complicated as the text gets longer. Long sequences turn the RNN into a very deep mesh, and it can take endless time to train, or the training may be unstable (Géron, 2019). Ba et al. (2016) employed layer normalization instead of batch normalization. Layer normalized RNN ensures that all collected inputs are fixed to a layer without a tendency to grow or shrink. Thus, a much more stable model can be created with the elimination of factors that cause vanishing or exploding gradients (Ba et al, 2016).

Another area of improvement for RNNs is the gradual forgetting of initial inputs when considering a long sequence. As a solution to this problem, Hochreiter and Schmidhuber (1997) presented a gradient-based method with Long Short-Term Memory (LSTM). LSTM gives faster results than the RNN algorithm when working with long text data and has been shown to provide better results in complex natural language processing tasks. Thus, any information in the array can be transferred with a line to reach where it is needed. LSTM makes a difference since the information is carried on a line without being transferred to the next section and is available where it is needed. This advantage can be used in a text as follows.

Ürünü dün sipariş ettim. Onu bugün teslim aldım.

*(I ordered **the product** yesterday. I received **it** today.)*

Pronouns can be regarded as an excellent example of this situation. The word “o” (it) in the second sentence refers to the product in the first sentence. Therefore, we expect the model to associate the word “the product” in the first sentence with the received item mentioned in the second sentence. This meaning should be carried in the series and associated with the related word in the text.

Although better results are obtained for RNN problems, it cannot be claimed that LSTM can provide the exact solution. These problems are associated with the model's

inability to transfer relevant context to remote locations in the array for long texts. For instance, the words in a customer review are sequentially considered as inputs to the model. A name that can be referenced as a definition next or a few sentences later may have disappeared entirely in the transfer process. It is a severe problem in NLP applications, and a number of studies are to solve it. RNNs and LSTMs process the texts token by token due to their solution architecture. As a result, it will not be easy to process the inputs simultaneously in the model.

2.1.4 Transfer learning

A model can be developed using the lower layers of a previously-trained neural network to solve another similar problem such that another problem's know-how can be used and transferred to the current solution. Thus, better results can be obtained with less training data after overcoming the problem of insufficient training data, which is especially common in natural language processing problems (Géron, 2019). Pre-trained layers can be incorporated into the model making various corrections at this stage. However, researchers should be aware that a preprocessing step may be required to standardize the inputs.

2.1.5 Attention mechanism

The vast majority of developments in NLP are based on the advances in deep learning research. Deep learning algorithms such as RNN and LSTM have achieved more effective results in NLP problems. However, there is still room for development in NLP. A technique that will give effective results, especially on texts, has been developed. It provides a solution to the crucial problems experienced by other deep learning methods, allowing the model to focus its attention on certain words (Vaswani et al, 2017). Attention mechanisms have significantly improved machine learning methods using neural networks in sentences exceeding 30 terms (Géron, 2019). Since texts used in NLP problems in real-life applications are generally longer than this value, attention mechanisms can be said to find a solution to the severe performance gap.

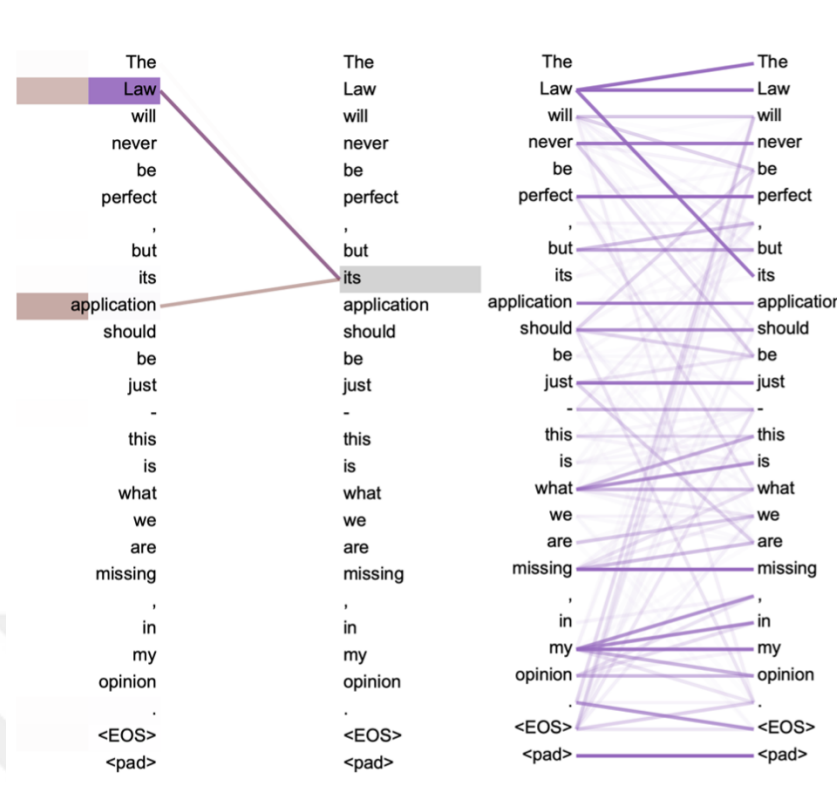


Figure 2.1 : An outlook of the attention mechanism for a specific word (Vaswani et al, 2017).

When writing a review on a product on an e-commerce site where we have purchased the product, we are driven by the features. For instance, if we have had a good or bad experience with the delivery of the product, we reserve extra room for the cargo operation in our comment. Similarly, if we had a problem with the supplier after we have received the product, we provide information in the review we are adding to the site especially on whether the seller has helped us in the post-purchase period. Likewise, attention mechanism is a technique developed to capture the targets that need attention in the text, as seen in Figure 2.1. In this mechanism, neural networks try to achieve this task focusing on some part of a subset of the information given to them.

2.1.6 Transformers architecture

In recent years, deep learning models such as RNN, CNN, and LSTM, which are frequently used in NLP, have been replaced by solutions that pre-train universal language models using semi-supervised learning and then customize these models with various architectural changes according to the problem (Géron, 2019). A new architectural approach, called Transformers, has been developed to perform multiple

tasks such as speech recognition, machine translation and text-to-speech transformation, which converts an input array into an output array, as indicated in Figure 2.2.

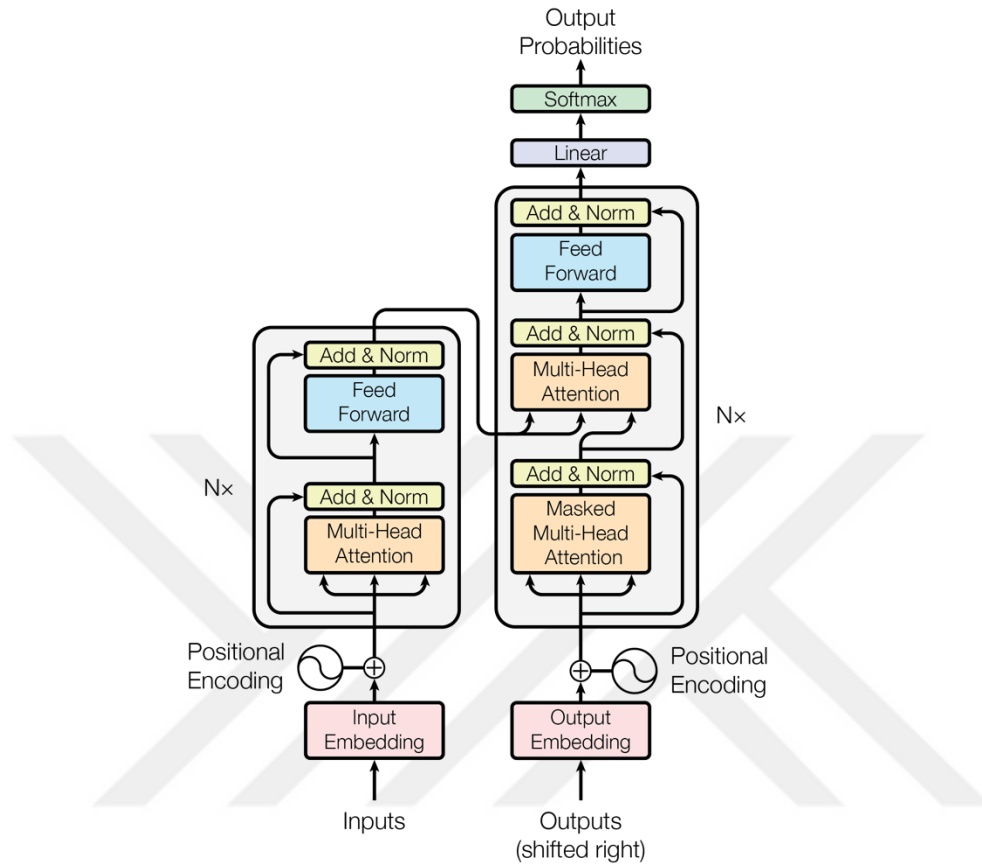


Figure 2.2 : Transformer - Model Architecture (Vaswani et al, 2017).

A transformer uses attention mechanisms when transmitting a whole array at once to produce an output rather than sequentially transferring all of the inputs. These attention mechanisms are implemented using the encoder-decoder architectural model. At each step, the memory cell in the decoder uses the weighted sum of all encoder outputs. Thus, the words the attention mechanism will focus on are determined. Unlike other deep learning methods, the transformers can be called an encoder-decoder architecture that uses attention mechanisms to transmit a complete picture of the entire array to the decoder at once, not sequentially.

2.1.6.1 Encoder

Text data are converted into word vectors in the embedding area. These vectorized text data are then inputted to the encoder block. As the first step in the coder block, the self-attention mechanism focuses on the relevance of words to the others in the text.

An attention vector is obtained for each word, showing the relationship with others in the text. The calculations are made taking the weighted averages to avoid focusing the attention on the analyzed word representation instead of the other word representations.

At the second step, a simple feed-forward neural network converts the attention vectors into a form that can be accepted by the next encoder or decoder layer. Unlike RNN, since each of the attention vectors is independent of each other, the attention vectors can be accepted as parallel inputs to the model simultaneously. Parallel processing can be performed on the GPU, and the processing process can be completed in a shorter time using higher efficiency technical infrastructure. This issue is one of the most critical points where Bidirectional Encoder Representations from Transformers (BERT) differs from other deep learning methods. All words can start and complete their operations in the encoder block one at a time, so the generation of encoded vectors for words can be performed simultaneously.

2.1.6.2 Decoder

The model must be trained with the actual values to classify the inputs. The decoder block has an embedding layer similar to the encoder block. Nevertheless, what goes into the embedding layer here are our target outputs, not our inputs. The model learns to what extent each word is related to others in the same array to capture the semantic context between the outputs. As in the encoder block, attention vectors are produced in the self-attention layer of the decoder block.

Unlike the encoder, the model first predicts using the previous results and compares the predictions with the actual values. As a result of this comparison, it updates the matrix values and learns through iterations. As another point of distinction with the encoder block, the model masks the outputs and provides predictions as a result of learning. The output value is masked, allowing a learning step between the model's predicted and actual values. These values are masked with 0 in the matrix in parallel operations. Thus, the attention network does not use the words after the target output.

The multi-head attention section is fed with the attention vectors produced in the first stage and the results from the encoder. This block reveals the relationship between the inputs and the outputs, and matching is done in this part. The vector generated for each word shows its relationship with all words, including the input and the output. Finally,

the attention vectors are transferred to the feed-forward layer. The conversion process is completed so that the output vectors can be processed by another decoder block or a linear layer.

2.1.6.3 Linear and softmax layer

Each element in the array is processed independently in another feed-forward layer, which is called the linear layer. The softmax layer converts the input into a human-interpretable probability distribution. The output is produced according to the highest probabilities.

2.2 Marketing Analytics

The study of data and modeling methods used to manage marketing resources and the performance monitoring of marketing activities is called *marketing analytics*. The analytics could assist the business decision-makers in understanding what drives the customer behavior, enhance the marketing strategies, and maximize the return on investment (ROI) by applying methodological and analytical processes to marketing-related data. Businesses can use marketing analytics to manage marketing resources more efficiently and retain profitable customers. Marketing analytics may also assist in collecting and analyzing data on customer brand preferences, shopping frequency, and purchasing trends (Iacobucci et al, 2019).

2.2.1 Market share

Market share, one of the metrics commonly used to measure marketing performance, is defined as the share of a business' sales relative to the sales of all firms in a market (Kumar and Reinartz, 2018). The market share can be calculated as a value over total sales or as a percentage represented by the firm's volume in the market.

$$\text{Market share of firm } j \text{ (\%)} = 100 \times \left[\frac{S_j}{\sum_{i=1}^I S_i} \right] \quad (2.6)$$

It is easy to obtain the sales performance of a business from its internal records. However, it may not be easy to learn the sales performance of its competitors in the market. Due to these difficulties, statistical methods are often used, and a sample is analyzed and then the results are generalized to the whole market. In addition, various market research reports are published, or further studies can be conducted for

competitor analysis. However, these resources can be either too general or too expensive when making a strategic decision. For this reason, businesses have started to follow alternative metrics that are highly correlated with market share.

Market share does not provide an insight into how much a customer has spent. Instead, it shows the sales of the businesses on a product basis and represents only a general performance aggregated over a category. Nowadays, marketing activities focus on the customer for a more personalized experience, and market metrics are defined by examining the customer's purchasing habits.

2.2.2 Share of requirements

Share of Category Requirement (SCR) corresponds to the ratio of the total sales of a seller to the sales made in the entire relevant category in the market. A more general SCR can be calculated when purchase data is unavailable at the individual level.

$$\text{Aggregated SCR (\%)} = \frac{\sum_{i=1}^I V_{ij_0}}{\sum_{i=1}^I \sum_{j=1}^J V_{ij}} \times 100 \quad (2.7)$$

Individual Category Requirement Share (iSCR) can be calculated for a specific customer when such detail is available. iSCR is found by dividing a seller's sales volume to a particular customer by the total category volume this customer spends on (Kumar and Reinartz, 2018).

$$\text{Individual SCR (\%)} = \frac{V_{i_0j_0}}{\sum_{j=1}^J V_{i_0j}} \times 100 \quad (2.8)$$

Thus, the metric shows to what extent an individual customer's category requirements is met by a seller. SCR is about customer loyalty like wallet share and is an essential metric for marketing units. While wallet share focuses on sales value, SCR focuses on sales volume.

2.2.3 Share of wallet

The wallet share is the percentage of a customer's purchase that is captured by the business selling the product. It can be measured at the individual (customer) level or at the aggregate level. Despite the wallet share is not a clear indication of the future revenues and profits that can be expected from the customer, businesses can learn how to plan their marketing strategy by learning from their customers' habits. Learning

about the share of wallet is comparatively more complicated than the share of category requirement. It gives the most accurate view of how a business can achieve its financial goals with the contribution of its customers.

The wallet share should be considered along with the wallet size. Businesses can make optimal decisions on resource allocation considering both of these dimensions. It is also substantial to define the difference between the market share and the wallet share. The calculation of the market share involves both buyers and non-buyers, while the wallet share is calculated based solely on genuine buyers. A firm's market share is the sum of the wallet shares of all its customers divided by the sales of all firms in the category in period t .

2.2.3.1 Individual wallet share

The individual wallet share is defined as the amount of money a customer spends in a category from a seller divided by the total amount this customer buys in that category from all sellers in the market. It indicates to what extent a customer's needs have been met by a seller in the category.

$$\text{Individual WS of firm } j_0 \text{ to customer } i \text{ (\%)} = \frac{S_{ij_0}}{\sum_{j=1}^J S_{ij}} \times 100 \quad (2.9)$$

2.2.3.2 Aggregate wallet share

The aggregate wallet share can be expressed as the seller's total revenue generated from all of the customers. It shows how much a seller has met its customers needs in the category it offers service.

$$\text{Aggregate WS of firm } j_0 \text{ (\%)} = \frac{\sum_{i=1}^I S_{ij_0}}{\sum_{j=1}^J \sum_{i=1}^I S_{ij}} \quad (2.10)$$

2.3 Related Studies

Research on Turkish sentiment analysis has generally been developed through traditional machine learning models. However, the transformers, which have been commonly used in recent years, have also achieved good results in text processing tasks. Thus, the number of studies using the Bidirectional Encoder Representations from Transformers (BERT) language model in Turkish sentiment analysis is increasing.

In addition to providing a review based on their experience, customers can also rate this experience that can be considered an advantage for researchers working in digital marketplaces. Besides, the marketplaces that design the channel through which they contact the customer in this way would eliminate a crucial natural language processing problem and facilitate the data acquisition process. This function of the digital marketplaces has made the use of scores in the annotation process widespread. Data annotation can be considered as one of the preparation tasks that need to be completed in natural language processing problems. Annotating customer reviews with ratings is a commonly used method (Acikalin et al, 2020; Sığircı et al, 2020).

2.3.1 Working areas

Acikalin et al. (2020) analyzed movie and hotel reviews using the multilingual model and the main model of BERT. The main model was used after translating the original data set into English to overcome difficulties encountered due to the limitations the Turkish language cause. The authors showed that the BERT model developed in the native language with the reviews translated from Turkish into English performed better than the multilingual model. In addition, the BERT was observed to be more successful than fastText (Bojanowski et al, 2017) when making predictions for a domain other than it was trained on.

One of the most fundamental NLP problems is the preparation of annotated training data. The annotation of data is an expensive effort-intensive process that requires high control. For this reason, Masarifoglu et al. (2021) focused on situations in banking and finance where annotated training data is limited or absent (Masarifoglu et al, 2021). First, the authors asked customers about NPS, which is used to measure their level of satisfaction with the available products. Then, they employed sample-free learning with a pre-trained XLM-Roberta-based model for Turkish sentiment analysis. This sentiment analysis study conducted in the banking sector showed that the model could make predictions for a domain other than it was trained on with sufficiently high accuracy. In addition, in cases where data annotation is not possible, it was observed that learning without the annotated training set gave better results compared to training a model with a different domain's data set.

Sigirci et al. (2020) used the Turkish BERT model for sentiment analysis of mobile applications in a marketplace. Text normalization – correction of spelling mistakes in

the reviews – was performed using Zemberek (Akın and Akın, 2007), an open-source library frequently used in natural language processing tasks in Turkish. Finally, Köksal and Özgür (2021) analyzed the sentiment of Turkish tweets about certain universities in Turkey using social media posts.

2.3.2 Data sets

Masarifoglu et al. (2021) used 6149 reviews to perform sentiment analysis with the responses of bank customers to NPS questionnaires. 65.7% of these reviews were positive, 31.8% were negative, and 2.5% were neutral. The reviews that were considered as neutral, in general, did not contain enough information to be labeled as positive or negative. Then, 60% of the data set was reserved for training the model, 20% for testing the model, and the rest for validating the model. Hence, the distribution of the labeled reviews was preserved for each sentiment class. In addition to the data set in the banking domain, which this study focused on, another data set containing reviews on universities was used.

Acikalin et al. (2020) used datasets containing hotel and movie reviews for sentiment analysis. Half of the data set, which had a total of 53,400 reviews, consisted of negative reviews, and the other half contained positive reviews. The second data set, which contained 11,600 hotel reviews, had the same split between negative and positive reviews like the previous data set.

Sigirci et al. (2020) used reviews written in Turkish in Google Play, a package manager and electronic media distribution/sales platform developed by Google for the Android operating system. Over two million customer reviews and ratings were collected, and BERTurk (Schweter, 2020), a Turkish BERT model, was employed. This native language model consists of 4.5 billion tokens, has a data volume of 35 GB and is case sensitive.

Köksal and Özgür (2021) used a dataset called BounTi, which included 7964 tweets on universities that were labeled as negative, neutral, and positive. The tweets were collected by searching for specific university names on Twitter. Similar tweets that had minor differences were removed using the Bigram Jaccard similarity. 72% of the observations were reserved for training, 8% for validation, and 20% for testing. In these sub-datasets, it was aimed to preserve the class distribution in the original

dataset. All of the sub-datasets thus included 29% positive, 53% neutral, and 18% negative tweets.

2.3.3 Data annotation methods

In Masarifoglu et al. (2021), annotating the reviews with a sentiment class was carried out independently by two experts working in the banking sector. A three-class (positive, neutral, and negative) classification was made. The labels these two domain experts provided were observed to be 96.09% compatible with each other. In addition, the BounTi dataset - which contained comments on universities and was manually annotated as positive, neutral, and negative - was used to test the performance of the current model.

In Acikalin et al. (2020), the movie reviews with a score of 1 or 2 were annotated as negative, those with a score of 4 or 5 were annotated as positive. In the other data set they used, the authors annotated the hotel reviews scored between 0-40 as negative and 80-100 as positive. Half of both datasets were used for training and the other half for testing. Therefore, a two-class classification was carried out.

Sigirci et al. (2020) study used almost two million reviews with ratings ranging from 1 to 5. The reviews given 1 or 2 stars were annotated as negative, whereas those with 4 or 5 stars as positive in the first model. In the second model, each rating was considered as a separate class, so a 5-class classification problem was developed.

In Köksal and Özgür (2021), two separate juries created the BounTi dataset manually. The juries first tried to identify the author's feelings about the university as positive, negative, or neutral during the annotation process. Then, they evaluated the tweets that had not been assigned to the same category. The authors carried out their study also with the characteristic features of the data source that could affect the model's success such as typos in the social media texts, including emojis and slang.

In the data sets recently used for sentiment analysis, observations created by users for shopping or review purposes were collected in various domains such as e-commerce, online media, or finance. Online channels where users submit a review and provide ratings were used for annotation. For sentiment analysis, datasets were created by automatically matching user ratings with the corresponding reviews obtained from these websites. This data handling approach is similarly used both in the English and Turkish languages. However, the practice was criticized that it might not be a

successful method for determining the sentiments of reviews from user scores automatically (Köksal and Özgür, 2021).

2.3.4 Classification algorithms

Masarifoglu et al. (2021) compared the performance of BERT-based models for sentiment analysis in banking with traditional machine learning methods.

Acikalin et al. (2020) used multilingual BERT along with BERT's main model that incorporated translations. Similarly, they used fastText in Turkish, and compared its results with those of fastText in English considering translations. fastText is a library developed by Facebook to learn word embeddings and text classification (Bojanowski et al, 2017). The authors also used two different benchmark models. One of them was a hybrid method developed using dictionary-based approaches and machine learning algorithms (Erşahin et al, 2019), whereas the other was a dictionary-based model built on an automatic translation from English to Turkish (Ucan et al, 2016). They asserted that BERT generated better results for the case of movie reviews. On the other hand, the hybrid model provided better results in the case of hotel reviews due to lack of sufficient volume for BERT to perform well. In addition, the authors showed that when there was sufficient data, better results can be obtained with the main BERT model working with translations from Turkish into English.

Sigirci et al. (2020) examined two-class and five-class classification models. The authors reported that the increase in the percentage allocated for training in the data set enhanced the model performance. A similar conclusion was made regarding the number of epochs. As the number of epochs increased, the model performance increased. However, the increments were not found to be linear. Depending on the expansion of the training data, the working time of the model also improved. Thus, the optimal point could be found. Finally, the effect of increasing the number of epochs in the training data differed among different sizes.

Köksal et al. (2021) analyzed the performance of multilingual and Turkish transformers such as multilingual BERT, XLM-Roberta, and BERTurk. The authors also made a social media analysis for a food brand using BERTurk, which was decided to be the best model. Besides, they observed that models trained with BounTi returned consistent results in other domains.

3. RESEARCH DESIGN AND METHODOLOGY

3.1 Research Objectives

This study aims to review the NLP and marketing literature and demonstrate how to use a hybrid methodology based on sentiment analysis and wallet allocation rule to gain market insights. It essentially focuses on sentiment classification and share of wallet estimation. In sentiment analysis, high performance is aimed at the predictions that will create the input for the wallet share using the transformer-based machine learning technique for natural language processing applications. On the other hand, the recommendations for future work may provide a predictable perspective on areas for improvement.

3.2 Research Model

This study adopts a hybrid approach, an end-to-end solution that combines two methods. As an NLP application, sentiment analysis serves as a prediction tool to identify customer satisfaction. At the same time, the wallet share rule is positioned as a tool where sellers can gain insights into their positions in the market based on the model results. The prediction model developed using attention mechanisms for sentiment classification constitutes the first analysis part of the study. In the second part, a second model developed for the wallet share rule calculates and estimates the wallet share those sellers have secured in the market for a specific category.

3.3 Sentiment Classifier – BERT

BERT (Devlin et al, 2019), developed by Google researchers, uses the transformers architecture which has recently shown to generate good results among natural language processing solutions using attention mechanisms. This model adopts pre-trained language representations to learn a general-purpose model of "language understanding" in a text. Therefore, the BERT model can be employed in various

natural language processing tasks such as sentiment analysis and text classification, as shown in Figure 3.1. BERT, which can be fine-tuned according to the business problem, is pre-trained on a vast corpus and deep bidirectional representations from unlabeled text with its scalable structure.

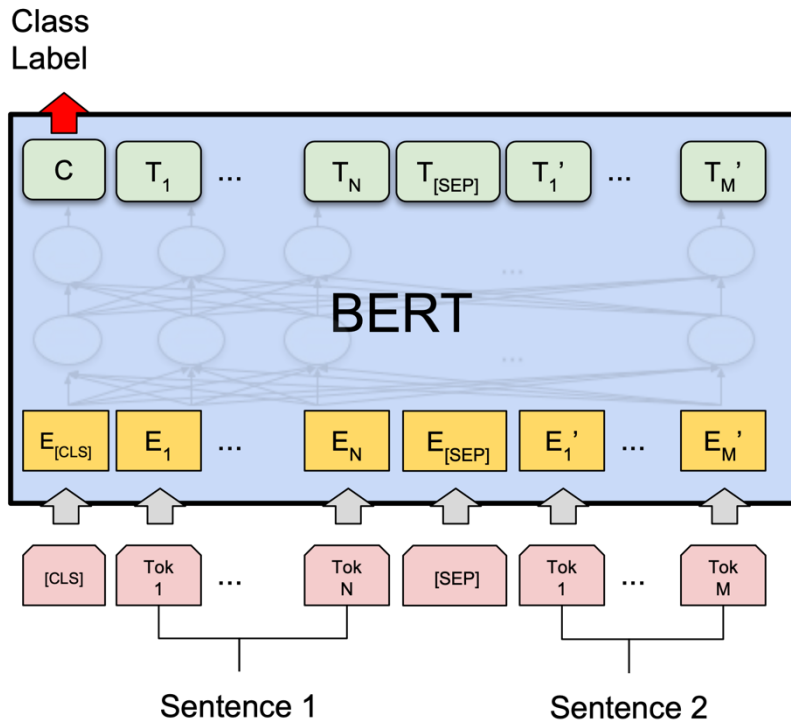


Figure 3.1 : BERT's Solution Architecture for Text Classification (Devlin et al, 2019)

The BERT model is pre-trained on a dataset from which it can learn the general language. On the other hand, customer reviews made in a digital marketplace operating in e-commerce may have some qualifications that are specific to its domain. Thus, fine-tuning assists in educating the model for domain-specific issues. The field in which the work is carried out has an effect on the NLP model. This two-step structure may apply to any NLP task that is executed similarly. For this reason, pre-training the BERT model on a corpus may serve as a preparation phase for the model to be specialized in the language understanding of the desired domain (Sun et al, 2019).

3.3.1 Tokenization

Word embeddings are substantial as they are the numerical representations of a text. Texts need to be tokenized before being used in the pre-training tasks: masked language model and predicting the next sentence. In the BERT model, preparing the

texts to create inputs to the model with their numerical representations is accomplished by embedding WordPiece. In this way, the BERT algorithm can understand different linguistic properties such as polysemy, where the word can have the exact spelling but more than one meaning. It can also make sense of word groups based on the context of the text using more accurate representations.

Out-of-vocabulary (OOV) are expressions with a dynamic structure, unlike the words in the fixed dictionaries. These expressions can often appear in texts such as customer reviews that are written in the language actively used in daily life. Using sub-word units, the detection problem created by OOV words can be solved, and the model results can be improved. It was developed by the Google Speech Recognition system for Asian languages such as Japanese and Korean. These languages have extended expressions with a wide variety of characters included in the text. Since the space as a character appears less often between words compared to other languages, the deterministic segmentation of terms in the text is needed (Wu et al, 2016). WordPiece notation was created to automatically learn the word units from large amounts of data. This technique for OOVs is used in BERT. Besides that, OOVs are ignored in most of the word embedding methods such as word2vec, which is the earliest word embedding model using neural networks that was developed by Google (Mikolovet et al, 2013), and GloVe, which was developed by Stanford University's researchers trying to capture the meaning of words according to their order in the text (Pennington et al, 2014).

Text data can be divided into smaller word groups using a trained WordPiece model. First, the borders of the words in the text are determined. Then, special characters are used to make the model recognize these boundaries. It provides to identify which token is used due to the spaces or division into subgroups (Wu et al, 2016). So, while a text is tokenized, they are divided into word subgroups. For example, a customer review to be tokenized at the document level can contain multiple sentences. The [CLS] token shows the beginning of the input. And the [SEP] token is appended to the sentences to distinguish multiple sentences in the text from each other. After adding the [CLS] and [SEP] tokens, tokenization is performed by assigning a numeric ID to words expressed with letters. Three types of embeddings are created to represent the entries. These are token embedding that represents the words, segment embedding that shows which

token belongs to which sentence, and position embedding that indicates the token location in the text.

Once the tokenized input is provided to the BERT model, it is expressed by the embedding vector that is represented by tokens, positions, and segments. These three types of vectors are collected and merged as an input to the encoder. The basic model of BERT consists of 12 encoders, whereas there are 24 of them in the large model. A vector, which is derived from word, position, and segment representations, is generated for each input that is considered an input to the encoders in the model. Moreover, it is ensured that the input and output dimensions are equalized (Sığircı et al, 2020).

3.3.2 Modelling

There are two task groups in the BERT model: pre-training and fine-tuning. The learning process takes place in the pre-training section, where the general language understanding is learned, and in the fine-tuning section, where the hyperparameters are tuned.

3.3.2.1 Pre-Training tasks

The BERT model performs unsupervised learning with unlabeled data during the pre-training phase. There are two pre-training tasks: the masked language model (MDM) and the next sentence prediction (NSP). In MDM, a text is first introduced to the model and some of its words are masked. Then, 15% of the words in the text are chosen randomly and hidden by tokenization as [MASK] in the masked language model. Subsequently, the model tries to predict the masked words. Learning the language patterns and filling the correct word in the blank are expected from the model. To succeed, it needs to capture the context of the text.

To train the native language model during the pre-training phase, BERT creates a dictionary from the dataset provided. Tokenized entries in this dictionary are processed through a frequency-based method. In NSP, for example, two sentences are used as an input when educating the main model trained to learn the language structure. The sentence is given as an input to the model in the following format:

[CLS] + Sentence 1 + [SEP] + Sentence 2 + [SEP]

The first sentence is introduced to the model using the [CLS] token to teach whether the second sentence can follow. The second sentence comes after the [SEP] token, followed by [SEP]. Next, the following sentence tries to predict. The model is trained to predict whether two sentences are related. For instance, a noun in a sentence may refer to a pronoun used in the following sentence. Here, there is a linguistic relationship between the sentences, and it is expected that the model could learn this problematic task.

3.3.2.2 Fine-Tuning tasks

The BERT model can independently perform encode text pairs and bidirectional cross attention tasks with its self-attention mechanism. Compared with pre-training, fine-tuning requires fewer resources and works faster. Inputs and outputs are explicitly given to the model. Fine-tuning aims to find the most suitable hyperparameters for the model. Both pre-training and fine-tuning adopt the same architectures except for the output layers (Devlin et al, 2019). The output corresponding to the first token [CLS] is used as an assigned class in text classification problems, as seen in Figure 3.1. The BERT model can also operate without the fine-tuning step. For instance, after pre-training, two different models, one feeding the linear regression model and the other one feeding the fine-tuned model, are compared in the sentiment analysis. BERT is a trained model for the target application that produces unsupervised features with a large and fixed model in pre-training. In fine-tuning, the parameters of the BERT model are optimized using the annotated data for a specific task (Souza and Souza Filho, 2022).

3.3.3 Performance evaluation

The BERT model performed well on the Stanford Question Answering Dataset (SQuAD) – a dataset for reading comprehension tasks consisting of more than 100,000 questions with answers to each question in the relevant text (Rajpurkar et al, 2016). This is one of the major reasons why researchers working in the field of natural language processing frequently use this model.

Generalization is one of the most critical objectives of NLP models. The model is required to give the best result for a data set it has not seen before. Therefore, it is critical for the model to perform sufficiently well without creating a serious problem such as over-fitting. As a result, a generalizable learning capability without over-fitting

or under-fitting is aimed in the model development process. The dataset is divided into training, validation, and testing subsets. The model is trained on the training dataset, and the best hyperparameters are searched using the validation dataset. Finally, the performance of the model is evaluated using a test dataset it has been introduced to for the first time.

3.4 Analyses

With the tool developed using the wallet share rule, and sentiment analysis using the BERT model, creating a hybrid solution is aimed within the scope of the study.

3.4.1 Sentiment analysis

Sentiment analysis is basically interested in extracting the opinions of customers contained in the reviews/statements and identifying their emotional tones as positive, negative, or neutral. Several studies have considered emotional tones such as joy, sadness, anger, and neutral (Huang et al, 2019). Sığircı et al. (2020) reported the successful implementations of sentiment analysis in various areas such as customer loss analysis, academic success estimation, election analysis, stock market expectations, and product review analysis. Furthermore, Özçift et al. (2021) compared the performance of BERT and ML algorithms in various Turkish NLP tasks such as cyberbullying detection, text classification, emotion recognition, and spam detection. Sentiment analysis has been a popular field of research nowadays; however, the expanding literature has mostly revolved around the English language and the number of studies working with the Turkish language is limited.

One of the problems that needs to be solved while working with the text data for sentiment analysis may occur when there are words that can have both a positive and a negative meaning (Sığircı et al, 2020). The following sentences can be considered as an example:

Sentence 1: “This sweater was too big for me.”

Sentence 2: “It has a big room where I put my things.”

The word “big” in these sentences may have a negative meaning when used for a sweater, but a positive meaning when used for a room in a house. Therefore, understanding a word's context in the sentence correctly and the domain of the data set

used for model training have an important role in achieving success while testing the model to overcome this problem in the sentiment analysis.

Since the data may have been obtained from various data sources for sentiment analysis such as a phone call speech converted into text, online consumer reviews, results of a survey conducted by a bank to evaluate the level of customer satisfaction, content shared on a specific topic in social media - it may contain various spelling mistakes and grammatical errors. The preprocessing stage thus plays a significant role in dealing with such issues in NLP problems.

3.4.1.1 Detail levels in sentiment analysis

Sentiment analysis is performed mostly at the document level to identify the emotional tone the customer reviews contain. It can also be performed at the word, sentence, syntactic structure, and aspect levels (Dehkharghani et al, 2017). In a document-level sentiment analysis, each document can be considered as an observation that has been assigned to a training set or a test set. Having more words in a classification at the document level compared to working at other levels allows more information for classification. Document level analysis could classify correctly with more inputs than the classifications at other levels.

Aggarwal (2018) showed how a sentence-level sentiment analysis with iterative neural networks classified into positive and negative sentiments and emphasized the importance of sequences. It was mentioned that the sequence-centered approach gave better results in making a more accurate classification. As in RNN, the BERT model works on the word sequence in sentiment classification problems and it can learn the contextual structure between words.

3.4.1.2 Challenges of Turkish

Turkish is an agglutinative language, and its words can take many inflexional and derivational affixes. It is pretty common to construct a Turkish word that corresponds to a full sentence in English (Dehkharghani et al, 2017). Therefore, it is more challenging to work with Turkish texts than with English texts, which the majority of studies on NLP have focused on. It is mainly due to the agglutinative structure of Turkish and the ability to derive words with different meanings using affixes that stem from the same root. Thus, it has been often stated that working with agglutinative

languages in NLP problems is more complicated compared to other languages (Sığircı et al, 2020).

Since language-specific features are used to analyze, in general, English texts, adapting these methods to another language such as Turkish is not an easy task (Acikalin et al, 2020). The rich morphological structure of Turkish and the unrestricted use of word ordering in sentences cause difficulties in analysis to be performed on texts. Acikalin et al. (2020) also emphasized that the relatively small number of natural language processing libraries developed with literature and technical studies in Turkish compared to other languages was one of the essential problems faced by researchers interested in this topic.

Intensive language preprocessing steps are required to model an NLP problem in a morphologically rich language like Turkish. While Turkish words consist of a root and inflectional suffixes that form the context, separate prepositions are used in English to create the similar context. Therefore, analytical languages like English can be analyzed and modeled more easily than morphologically rich languages like Turkish. However, the number of suffixes added to a Turkish word and their combinations constitute a significant problem in language analysis (Özçift et al, 2021).

3.4.2 Wallet allocation rule

Customer satisfaction related to the products and services that is provided by a seller is a critically important criterion for loyalty. On the other hand, the need of customers may be met by other sellers in the market as well. Besides, no meaningful relationship between customer satisfaction and wallet share has been discovered, as seen in Table 3.1. As Keiningham et al. (2015) claimed, the Wallet Allocation Rule shifted the focus of businesses from factors that drive customer satisfaction metrics such as customer satisfaction or net recommendation scores to factors that influence the order of preference of the brand in the eyes of customers (Keiningham et al, 2015). Therefore, the order in which customers prefer a seller among all category expenditures they use is essential. This order of preference can be easily calculated with the help of a simple formula called the wallet share.

Table 3.1 : Customer Satisfaction Levels and Share of Wallets (Keiningham et al, 2015).

Customer ID	Satisfaction	Share of Wallet
1	9	30%
2	8	60%
3	7	15%
4	9	75%
...	...	...
997	10	35%
998	9	45%
999	8	25%
1000	8	50%
R-Square		1.13%

Converting absolute satisfaction levels into relative ranks can account for more than 20% of the difference in customers' share of category expenditure. What counts is the consumer's perception of performance on the seller compared to others, not the rating given by the customer.

Keiningham (2015) claimed that converting customers' ratings for brands they use in order of preference would allow for creating a link between data from customer surveys and actual purchase patterns. Also, the rule could determine the wallet share with high precision by knowing the seller's priority order and the number of sellers the customer shopped at. Traditional customer metrics such as satisfaction and net recommendation score, which businesses nowadays monitor, can be reinterpreted with this method. Alternative metrics can also be tailored to be used as an input to the study's rule-based model, taking business processes and platform capabilities into account. These metrics should be analyzed using the wallet share rule within the context of tracking strategic goals and performance management.

Sellers should reshape their customers' satisfaction measurements since customers evaluate them compared to alternative competitors. The Wallet Allocation Rule is a measuring technique that can be used for such a purpose. A seller's market share and

its relative ranking among competitors are asserted to be linked. This rule is nevertheless a valuable statistic for sellers to comprehend their market share. The rule, maybe most critically, allows vendors to establish a correlation between customer delight and measurable financial results.

Sellers cannot realize their current status without addressing the competition when defining their market strategies, and they cannot design a feasible roadmap to achieve their objectives. While customer satisfaction differs minimally between the first and the second priority, the financial impact is significant. Regardless of whether the seller gets a good rating from a customer, the customer purchases from the highest rated seller, and the high rating received by the seller may not result in an active agreement.

Moreover, attaining the highest customer satisfaction score may not always be sufficient on its own. The number of sellers with the highest ratings also matters. For instance, the customer's wallet may be splitted among the sellers that have the highest ratings. Furthermore, the more sellers a customer purchases from in a category, the lower the potential wallet share of that customer for everyone. Therefore, sellers should know why their customers buy both from them and from other sellers in the market that provide similar services or products. Understanding the customer requirements, they may design a solid marketing strategy and prepare a robust action plan to be the first option of the customers in the market.

To calculate the wallet share of a seller from a particular customer, the order of preference of the seller and the total value of the sellers used by the customer are calculated through the wallet share rule as formulated below:

$$\text{Wallet Share} = \left(1 - \frac{PB}{TB + 1}\right) \times \left(\frac{2}{TB}\right) \quad (3.1)$$

To use the Wallet Allocation Rule, first, a customer research is conducted for the determination of sellers that customers regularly purchases from in the product category requested to be analyzed during data preparation. This research can be prepared to measure the level of customer satisfaction. During the data collection phase, a hybrid metric that is formed by one or more of the loyalty metrics indicating customer satisfaction is used for each seller in the market preferred by the customers. It is then converted into a preference order based on these values.

4. EXPERIMENTS

4.1 Sampling and Data Collection

Product reviews provided for one of the leading e-commerce marketplaces operating in Turkey are evaluated in this study. The data set contains several details about the review submitted by users. There are columns that indicate the category a product belongs to and the sellers offering this product in the marketplace. This dataset contains 1,048,575 reviews (CONTENTS) for a total of 3,183 product categories (CATEGORYNAME) and 39,722 sellers (SELLER_ID).

4.2 Determination of Product Categories

Category-based modeling aid in dealing with the obstacles related to domain-dependent problems. For instance, a word as a feature of a product category that contains a positive sentiment may be considered to contain a negative sentiment for another product category.

One of the biggest obstacles to overcome in implementing natural language models is the volume and quality of the training data. Therefore, the product category containing the most observations in the data set was chosen as the base (primary) category. The largest number of reviews in the data set are available for the category “case”.

In a report on e-marketplace platforms that was prepared by Turkish Competition Authority, the effect of COVID-19 on e-commerce was discussed (Sarıçiçek et al, 2022). The study tries to specify in which categories the pandemic caused a leap in e-commerce. The years 2019 and 2020 should be considered together to define the impact of COVID-19. The study reported that e-commerce volume increased by 64% in the first six months of 2020 compared to the same period in 2019. The rate of increase in 2020 is substantially higher than the ones recorded in previous years.

When category-based retail e-commerce sales in 2019 and 2020 are compared, a decrease of approximately 53% is observed for “Travel and Accommodation”. Besides that, an increase of 42% for “Fashion and Beauty” and an increase of over 40% for “Electronics and Media Products” are noticed.

The present study thus focuses on these two main categories that have been impacted by COVID-19 positively. Subsequently, product categories with the highest number of reviews in the data set are investigated. The product category "Men's Perfume" under the main category "Fashion and Beauty" and "Mobile Phone" under the main category "Electronic and Media Products" are selected eventually.

4.3 Data Annotation

Small amounts of training data are generated due to the difficulties in the manual annotation for text classification problems. However, the model's training process can be created with large data sets when the scores provided by customers are used in the annotation of reviews. Each customer review in the dataset has a rating from 1 to 5. The reviews are annotated by subtracting 1 from the ratings so that they can be taken as an input to the BERT model as shown in Table 4.1.

Table 4.1 : Customer Ratings and Corresponding Classes.

Ratings	Classes
1 star	0
2 stars	1
3 stars	2
4 stars	3
5 stars	4

4.4 Configuraitons

Datasets in real-life problems may not be ready-to-use in machine learning and natural language processing models. The data actively used in various business models should be extracted for the natural language processing problem. Before the preprocessing and modeling steps, some adjustments may be necessary to apply the models to the

extracted dataset. The configuration steps given below are followed to prepare the dataset for the preprocessing:

1. Reviews that are not null in not only the content, but also the product score, the category and the seller ID columns are filtered to create the final dataset. After data cleaning, 479,377 reviews (CONTENTS) for a totao of 3,042 product categories (CATEGORYNAME) and 31,893 sellers (SELLER_ID) left.
2. In the e-commerce marketplace, products can be rated from 1 to 5 by customers. The scoring should be revised to change between 0 and 4 so that they can be inputted to the BERT model in a similar order.
3. The revised data set is filtered for each category to be examined, and a sub-dataset belonging to a specific category is produced.
4. There should be enough reviews to measure how satisfied users are with the product/service offered by a seller. For this reason, sellers for which ten or more product reviews are available were determined, and the data set was filtered to include only these sellers.

4.5 Data Pre-processing

Structural difficulties are encountered in natural language processing solutions developed with agglutinative languages such as Turkish. Preprocessing is thus necessary for accurately modeling the data set due to several issues such as typos. The texts that will be inputted to the model are made appropriate and regular so that the model can provide a better classification at this step. Additionally, this step includes the most common techniques that are used while cleaning the data. The preprocessing steps that were followed to provide the model the inputs with better quality are listed below:

1. **Stop Words Removal:** The first preprocessing step is concerned with removing the stop words (the words in any language or corpus that occur frequently). Since stop words have no specific meaning in the texts, removing them will not affect the quality of the reviews. Since they carry no additional or valuable information for the model, redundant words are removed so that the model could focus on meaningful words for the classification. The stop words were removed using the NLTK library (Bird et al, 2009).

2. **Lowercase Conversion:** The preprocessing stage and the following stages in the NLP application can benefit from converting all texts to lowercase. “Product” and “product” are the same words; however, if they are not converted to the lower case, these two words may be represented as two different words in the vector space model since the first word begins with an upper case. The upper cases are changed with lower ones to avoid the unnecessary explosion in the number of dimensions in the input representation.
3. **Detecting and Removing URLs:** Data collected from digital marketplaces may contain URL links, which must be removed from to avoid ambiguity. The URL links are detected in the preprocessing phase and removed from the text.
4. **Replacing Extra Spaces:** Extra whitespaces and tabs do not bring any extra information in an NLP application. Singularization of them can provide much more basic structures for the modeling phase.
5. **Replacing Special Characters:** Non-word characters may be found in comments, references, currency numbers, and of course, product reviews. These characters do not benefit text comprehension and cause noise for NLP pipelines. As a result, they can be replaced with white spaces.
6. **Replacing Digit Characters:** Text, including digits, can be difficult for NLP models to learn. Since this scenario is hard to analyze, digits can be eliminated. As a result, they are replaced with an empty string to generate regular expressions.
7. **Replacing New Lines and Tabs:** Numerous new lines and tabs may appear for no apparent purpose in a customer review dataset. When collecting data from an e-commerce marketplace, newlines and tabs that are necessary for the platform to display the content are not needed in the dataset. Besides, they are converted into meaningless characters and can appear in the dataset as \n or \t. As a result, new lines and tabs can be also replaced with a space.
8. **Removing Punctuations:** The removal of punctuations is another text processing technique. Any punctuation in a text is changed as a white space using a string module. Below is a list of the 32 major punctuation marks that can be avoided:

'!"#\$%&'()*+,-./:;<=>?@[\\]^_`{|}~'

9. Removing Specific Strings: A piece of text which is unrelated to the context of the customer review dataset and the final results can be expressed as noise. If not required, it could be helpful to eliminate specific strings to clear the dataset from these noises.

4.6 Splitting the Dataset

One of the best ways to see how well an NLP model will perform with new samples is to monitor the model's performance in the production environment. On the other hand, testing NLP applications with users extremely risky. The application should be tested and taken to the production environment after its performance has been shown to be adequate. For this reason, it is necessary to divide the customer review dataset into two groups as the training dataset and the test dataset.

While performing a sentiment analysis, the training dataset is used to train the model. The test dataset is used to test the model and observe how it performs with examples that it has not seen before. Thus, the error rate in new samples is estimated using the test data set. The issue that needs to be checked here is whether the model is overfitting to the training dataset. Even though it has a relatively low error rate on the training data set, there might be an overfitting problem if the model has produced higher error rates using new observations.

30% of the data was reserved for testing, and the remaining part was allocated for training to evaluate the classifier's performance. Before the split, the data set could be shuffled. Randomization could be controlled with an integer parameter in the Scikit-learn module (Pedregosa et al, 2011). This parameter ensures that the same results are obtained across different calls. One of the most commonly used integer random seeds (42) was preferred in our model.

Moreover, text classification problems might include a substantial imbalance in the distribution of the target class. For instance, %90 of the samples includes reviews with 5 stars, whereas only %1 of the reviews were given 1 star. In such scenarios, stratified sampling should be adopted to provide that relative class frequencies in each subset (training and test) are sustained. Each set has the same number of samples from each target class. Thus, the data set allocated for testing becomes representative of the real-life application. The stratified element preserves the class ratios for both the training

and test datasets. The stratified element is used on a seller basis to separate customer comments in the same class ratios for both of these subsets.

4.7 Tokenization, Pre-Training and Fine Tuning

Customer reviews are classified into multiple classes using the BERT algorithm. The final data set is summarized in two columns named as text and class. The former is basically a string variable that consists of the customer reviews and the latter is an integer variable containing the customer ratings. The pre-trained model with a size of approximately 35 GB, dbmdz/bert-base-turkish-uncased (Hugging Face, 2022), was used. Because the device used to train the model has no GPU, only CPU is available, and CUDA configuration is not used.

The ratings of customer reviews were incorporated into the models in four ways. The first approach treated each rating as a distinct class and a five-class model was built. However, as the number of classes increases, the model's performance decreases. Thus, alternative ways were sought. The ratings of customer reviews were grouped within themselves. In the second approach, customer reviews with 1 or 2 stars were labeled as negative, those with 4 or 5 stars as positive, and the rest as neutral. This is a three-class classification model. A two-class classification model was built in the third approach with a slight modification of the three-class model: the customer reviews with 3 stars were excluded, keeping everything else the same. Finally, another two-class model was built labeling 1-star customer reviews as negative and 5-star customer reviews as positive, while discarding any review with 2, 3, or 4 stars.

Moreover, transformers need to make argument arrangements for the type of the problem and the dataset. BERT has a number of configuration settings that allow users to customize the model with respect to the needs of their problem. Thus, the focus should always be on the data, and its quality and consistency. Understanding the data is essential to find the proper arguments. Thus, researchers should first try to understand the problem and the data, and then find the correct arguments. Libraries that are easy to use contain the typical configuration for all tasks. The number of epochs the model will be trained on can be configured, and the model's epoch number is set at 2. At the end of each epoch, a model checkpoint is saved. Even if a cached file of the input data exists in the cache directory, the input data is reprocessed. The trained

model is stored in the output directory, overwriting any previously stored models at the exact location.

4.8 Making Predictions

The prediction is made using the observations from the actual test data that the model did not see before. The predictions generated by the text classification model on the test data were compared with the ratings to reveal the model's performance. The model tries to classify the customer reviews according to ratings. The encoder component used by the BERT model takes the token array as an input after converting the tokens into vectors. The product categories selected, and the corresponding number of classes were applied in the prediction part of the model. The effect of these categories and the number of classes on prediction success can be examined accordingly.

Table 4.2 : The Weighted-Averaged F1 Scores.

Category Name / Class Number	Case	Men's Perfume	Mobile Phone
5-Class	0.7232	0.7337	0.7618
3-Class	0.8631	0.8626	0.9016
2-Class – Type 1	0.9300	0.9219	0.9395
2-Class – Type 2	0.9550	0.9440	0.9663

Similar effects were observed in the three product categories when the data annotation process was carried out differently. When the model was requested to make more detailed predictions, the incorrect predictions increases. Besides, the entire data set was considered by the five-class and three-class models. 3-star product reviews were not incorporated into the first version of the two-class classification model. Likewise, the reviews with 2, 3, or 4 stars were not considered by the second version of the two-class classification.

Evaluation was made by taking the weighted average f1 scores for each class and each product category as shown in Table 4.2. Both of the two-class models generated the best results across all of the product categories analyzed. However, it should be noted that these models were not built on the complete data since some of the reviews were removed from the original dataset that has been received from the digital marketplace. In other words, some of the real-life observations were ignored when solving the

problem. Regarding the three-class and five-class models, in which all of the observations were included, the weighted averages of f1 scores for the three-class indicate a better performance. The data annotation using three classes appears to be the ideal case considering the inclusiveness and model performance.

4.9 Wallet Share

A significant increase in revenue can be observed when factors supporting customer satisfaction and loyalty and contributing to getting ahead of the competition are emphasized (Keiningham, 2015). Focusing on the competitive position of the product/brand/firm that is examined with the wallet share rule in a digital marketplace, inferences can be made on potential financial impacts.

1. For the three product categories determined (i.e., case, men's perfume, and mobile phone), customer reviews in a digital marketplace were collected. The customers provided the seller information in their reviews on the products . Thus, which customer review relates to which seller's product can be determined. The reviews of customers who have purchased a product have been evaluated.
2. A customer satisfaction metric was formed by collecting the reviews on the three products in the digital marketplace. Since there was no information about which customer reviewed the product in the existing dataset, reviews were assigned to random customers. When the effect of number of customers was analyzed, distributing all of the reviews among more people was observed to increase the model's success rate. Thus, the case with the highest number of customers was selected. In the proposed solution, the wallet share was calculated for a total of 500 customers as shown in Table 4.3. Other alternatives involving different class sizes were also calculated, but the conclusion did not differ. Much more details can be found in Appendix A.
3. The sellers and their rankings were determined to calculate the wallet share of a seller from a customer. While generating this ranking, the ratings estimated

through customer reviews were used. Finally, the wallet share was estimated by incorporating these values into the function developed.

4. Iterative functions were developed for each customer and each seller. First, each customer's wallet share ratings were averaged to estimate the wallet share those customers allocated to a particular seller. The wallet share calculated in this study analyzes the share of the sellers based on the expenditures of customers shopping in this marketplace as shown in Table 4.4. Other alternatives as discussed earlier (different class sizes and customer volumes) were calculated. Related details can be found in Appendix B.

The proposed solution focuses on the average wallet share allocated to a seller by all customers who shop in the digital marketplace. Thus, a general idea that is more comprehensive about the financial benefit of a seller from this marketplace could be obtained.

Table 4.3 : Wallet Share Performance Metrics (Three-Class Approach).

Category/Number of Customer	Case			Men's Perfume			Mobile Phone		
	100 Cus	300 Cus	500 Cus	100 Cus	300 Cus	500 Cus	100 Cus	300 Cus	500 Cus
MAE	0.1164	0.0388	0.0232	0.1982	0.0660	0.0396	0.1278	0.0426	0.0255
R2	0.9983	0.9983	0.9983	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
MAPE	0.0848	0.0848	0.0848	0.0368	0.0368	0.0368	0.0385	0.0385	0.0385
MSE	0.0358	0.0039	0.0014	0.0756	0.0084	0.0030	0.0353	0.0039	0.0014

Table 4.4 : Comparison of the Actual and Predicted Wallet Shares (Three-Class Approach with 500 Customers).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	1880900	7.2235	7.3082	2107203	26.4442	26.2790	329366	18.2066	18.1266
2	376354	6.8035	6.8195	2018650	10.9109	10.8090	32315	6.6533	6.5633
3	759501	5.4947	5.6836	2037808	7.3114	7.3314	920478	5.1666	5.1166
4	612418	3.3826	3.4300	1747712	4.3961	4.4961	42154	3.0433	3.0233
5	2343193	2.6128	2.6212	1658884	3.5647	3.4928	2809218	2.4700	2.4600
6	104357	2.6126	2.8052	732566	3.1647	3.2514	728461	2.1666	2.1500
7	2411290	2.6025	2.6049	364365	3.1614	3.1847	515457	1.9700	1.9600
8	1696631	2.5844	2.4057	3401700	2.8666	2.7933	995645	1.9666	2.0000
9	899464	2.3700	2.2422	34142	2.3200	2.3566	2374001	1.9166	1.8666
10	618452	1.9328	2.0084	431387	2.1766	2.2433	1672829	1.5200	1.5166



5. FUTURE WORK & CONCLUSION

5.1 Future Work

A hybrid solution involving sentiment analysis and market analysis was proposed in this study. However, better results can be achieved by improving the approaches adopted in both of the analyses. On the other hand, while performing the results of the present study, it should be ensured that this study is a hybrid method and that the improvement made in a part of it contributes to the actual output. In this context, development steps that can be used in future studies are mentioned in this section.

5.1.1 Model retraining

After the model selection is made, the analysis service receives the latest updates and continues to meet the incoming requests. One of the most vital aspects of this system is that the improvement works are automated while the model is running actively. The literature has generally focused on the model development process in NLP studies. However, studies on the sustainability of the model are limited. Retraining mechanisms may not be implemented easily where it is impossible to instantly annotate texts with user ratings as in the case of online customer reviews. In the improvement step of the model, where data cannot be instantly annotated, an effort for manual data annotation will be required and automation will be limited. Nevertheless, such a continuous improvement step can be easily added with online reviews and ratings from a digital marketplace.

Application architecture, in which the first model trained can evolve during system operation through an integrated incremental learning mechanism and feedback, can be also constructed. The retraining mechanism has been shown to improve classification accuracy without degrading overall system performance (Capuano et al, 2020). However, retrained models may not always perform optimally. Hence, a threshold can

be determined and the retrained model can replace the existing model once a predetermined level of success has been attained.

In the first step, success metrics for the model are defined with the help of customer needs, business problems, and experience gained in similar projects. Next, model performance is evaluated using these metrics. If the model performance can be improved using the retrained model, the new model is deployed to the system and actively used. On the other hand, if the retrained model cannot meet the necessary system requirements and cannot improve the prespecified metrics compared to the current model, the current model can continue its life cycle.

5.1.2 Aspect-based sentiment analysis

In sentiment analysis, some expressions to be evaluated may be interpreted in more than one way. For instance, a particular feature that was mentioned negatively in a statement could have a positive meaning in another one. Aspect-based sentiment analysis can be considered a multi-class classification problem. Aspect and sentiment classification are two primary tasks for a given expression.

In aspect-based sentiment analysis, aspect extraction and sentiment analysis can be carried out as two separate tasks (Xiao et al, 2021). Pipeline models may be created by using these two tasks sequentially. Also, there is a combined model that classifies both the sentiments and the aspects as labels. Lastly, there is a model that produces results in distinct layers but in which the whole model can be optimized in the same framework.

There have been several studies on sentiment analysis that used predefined categories. In a study concerning restaurant reviews, aspect extraction, the first task of aspect-based sentiment analysis, was performed with RNN-based aspect tagging using five predefined aspects: Service, Ambiance, Price, Location, and Food (Dhanush et al, 2016). Another study aimed at identifying four predefined categories. A group of five words were selected to represent each category. The similarity between a sentence and a category was calculated through "Soft Cosine Similarity Measure" that considered the sentence and the manually selected set of words corresponding to that category (Ghadery et al, 2019).

To eliminate the human effort required in the data marking process, Kumar et al. (2021) considered the names in laptop and restaurant reviews as potential targets and

the adjectives or adverbs as potential expressions of emotion during the data annotation. The authors thus employed the POS labeling method to detect the nouns, adjectives, and adverbs in a sentence in the datasets (Kumar et al, 2021).

In studies on sentiment analysis, aspects could be found also using various extraction methods. Aspect extraction techniques can be examined in five main groups: language rule model, sequential model, topic model, deep learning, and hybrid methods (Madhoushi et al, 2019). Aspect extraction was observed to generate better results through supervised learning compared to rule-based ones. However, data annotation and the need for high-quality features bring difficulties in performing aspect extraction through supervised learning. Deep neural network (DNN) models using pre-trained word embeddings are also limited by their quality (Xiao et al, 2021).

5.2 Conclusions

Text classification has been widely applied in marketing activities by many of today's leading companies. Sentiment analysis is a common type of text classification that provides a class to a sequence of text such as positive, negative, or neutral. On the other side, the Wallet Allocation Rule could be summarized simply as a formula that combines the seller's rank and the number of sellers in a market to calculate the wallet shares. Combining these two methods, a tool was designed to enable sellers to realize market intelligence. The present study provides a systematic literature review on marketing and NLP and proposes a hybrid approach based on sentiment analysis and wallet allocation rule.

As an application of the research methodology, a case on how to pre-train and fine-tune BERT through customer reviews obtained from digital marketplaces to determine the degree of customer satisfaction was presented. Using the results of the prediction model, businesses can determine what it takes to be the first choice of the customers in the market by understanding why customers would prefer them or their competitors. With the model incorporating the wallet allocation rule, it was depicted how sellers can monitor the shares their customers allocate them in their wallets considering their own sales along with the sales of their competitors on an e-commerce platform.

The prediction model, which constitutes the first part of the hybrid approach, was developed with respect to the characteristics of the problem after reviewing the related

literature and examining the practical applications. After the completion of data analysis, annotation, preprocessing, and modeling stages, text classification was performed and prediction results were produced. Then, four alternative models that vary based on the number of classes they considered were evaluated, and the 3-class classification model was decided to be more practical and rational to proceed with. The percentage of correctly classified cases was found as 0.8631 for the “Case” category, 0.8626 for the “Men’s Perfume” category, and 0.9016 for the “Mobile Phones” category. Since the Wallet Allocation Rule calculates the customer-based wallet share, the customer reviews were randomly assigned to the available customers. Alternative models were created based on varying customer sizes (i.e., 100, 300, and 500) to detect whether the number of customers has an effect on estimating the wallet share. Once the customers were created, the results were assigned to them randomly. The highest number of customers were found to result in the lowest prediction error. The wallet shares are calculated assuming that the online reviews collected for three categories are received from 500 customers.

The present study claims that businesses should analyze the sentiments of online customer reviews to gain a better understanding on how satisfied their current customers are. The results of such an analysis can also be used to compare a seller’s market performance against its competitors. Hence, this study provides a tool business can follow the trends and emerging topics in their market using these results. In addition, they can reshape their primary strategic business goal, financials, using the insights gained by the wallet share estimation.

REFERENCES

- Acikalin, U. U., Bardak, B., & Kutlu, M.** (2020, October). Turkish Sentiment Analysis Using BERT. *28th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4).
- Aggarwal, C. C.** (2018). *Neural Networks and Deep Learning*. Yorktown Heights, NY: Springer.
- Akin, A. A., & Akin, M. D.** (2007, October). Zemberek, an open source NLP framework for Turkic Languages. *Structure*, (pp. 1-5).
- Alamsyah, A., Rahmah, W., & Irawan, H.** (2015). Sentiment Analysis Based On Appraisal Theory For Marketing Intelligence In Indonesia's Mobile Phone Market. *Journal of Theoretical and Applied Information Technology*, (Vol. 82(2), pp. 335-340).
- Ba, J. L., Kiros, J. R., & Hinton, G. E.** (2016). Layer Normalization. *arXiv preprint arXiv:1607.06450*.
- Baroni, M., Dinu, G., & Kruszewski, G.** (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 238-247).
- Bird, S., Loper, E., & Klein, E.** (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.
- Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T.** (2017). Enriching Word Vectors with Subword Information. *Transactions of the association for computational linguistics*, (Vol. 5, pp. 135-146).
- Brown, P. F., deSouza, P. V., Mercer, R. L., Pietra, V. J., & Lai, J. C.** (1992). Class-Based n-gram Models of Natural Language. *Computational linguistics*, (Vol. 18(4), pp. 467-480).
- Capuano, N., Greco, L., Ritrovato, P., & Vento, M.** (2020). Sentiment analysis for customer relationship management: an incremental learning approach. *Applied Intelligence*, (Vol. 51(6), pp. 3339-3352).
- Chollet, F.** (2017). *Deep Learning with Python*. New York, NY: Manning Publications.
- Cooil, B., Keiningham, T., Aksoy, L., & Hsu, M.** (2007). A Longitudinal Analysis of Customer Satisfaction and Share of Wallet: Investigating the Moderating Effect of Customer Characteristics. *Journal of Marketing*, (Vol. 71(1), pp. 67-83).

- Dehkharghani, R., Yanikoglu, B., Saygın, Y., & Oflazer, K.** (2017). Sentiment analysis in Turkish at different granularity levels. *Natural Language Engineering*, 23(4), 535-559.
- Deliçay, M.** (2021). *Perakende E-ticaretin Yükselişi: Fırsatlar, Sorunlar ve Öneriler*. Türkiye Cumhuriyeti Cumhurbaşkanlığı Strateji ve Bütçe Başkanlığı.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K.** (2019, May). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv preprint arXiv:1810.04805*.
- Dhanush, D., Thakur, A. K., & Diwakar, N. P.** (2016, May). Aspect-based Sentiment Summarization with Deep Neural Networks. *International Journal of Engineering Research & Technology (IJERT)*, 5(05), 371-375.
- Erşahin, B., Aktaş, Ö., Kılınç, D., & Erşahin, M.** (2019). A hybrid sentiment analysis method for Turkish. *Turkish Journal of Electrical Engineering & Computer Sciences*, 27(3), 1780-1793.
- Géron, A.** (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition*. O'Reilly Media, Inc.
- Ghadery, E., Movahedi, S., & Faili, H.** (2019, June). An Unsupervised Approach for Aspect Category Detection Using Soft Cosine Similarity Measure. *arXiv preprint arXiv:1812.03361*.
- Giudice, P. L., Musarella, L., Sofo, G., & Ursino, D.** (2019). An approach to extracting complex knowledge patterns among concepts belonging to structured, semi-structured and unstructured sources in a data lake. *Information Sciences*, 478, 606–626.
- Göl, H., Ot, İ., Çakır, İ., & Sönmez, E.** (2022). *E-ticaretin öne çıkan başarıları, tüketici davranışlarında değişim ve dijitalleşme*. İstanbul: Deloitte Digital - TÜSİAD.
- Hochreiter, S., & Schmidhuber, J.** (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 735-1780.
- Huang, Y.-H., Lee, S.-R., Ma, M.-Y., Chen, Y.-H., Yu, Y.-W., & Chen, Y.-S.** (2019). EmotionX-IDEA: Emotion BERT – an Affectional Model for Conversation. *arXiv preprint arXiv:1908.06264*.
- Hugging Face.** (2022, 05 12). Retrieved from dbmdz / bert-base-turkish-uncased: <https://huggingface.co/dbmdz/bert-base-turkish-uncased>
- Iacobucci, D., Petrescu, M., Krishen, A., & Bendixen, M.** (2019). The state of marketing analytics in research and practice. *Journal of Marketing Analytics*, 7(3), 152-181.
- JPMorgan Chase & Co.** (2021). Retrieved from Global E-Commerce Trends Report: <https://www.jpmorgan.com/content/dam/jpm/treasury-services/documents/global-e-commerce-trends-report.pdf>
- Keiningham, T. L.** (2015). *The wallet allocation rule: Winning the battle for share*. John Wiley & Sons.
- Keiningham, T., Cooil, B., Aksoy, L., Andreassen, T., & Weiner, J.** (2007). The value of different customer satisfaction and loyalty metrics in predicting customer retention, recommendation, and share-of-wallet. *Managing service quality: An international Journal*.

- Köksal, A., & Özgür, A.** (2021, September). Twitter Dataset and Evaluation of Transformers for Turkish Sentiment Analysis. *29th Signal Processing and Communications Applications Conference (SIU)*, 1-4. IEEE.
- Kumar, A., Gupta, P., Balan, R., Neti, L. B., & Malapati, A.** (2021, July). BERT Based Semi-Supervised Hybrid Approach for Aspect and Sentiment Classification. *Neural Processing Letters*, 4207-4224.
- Kumar, V., & Reinartz, W.** (2018). *Customer Relationship Management: Concept, Strategy, and Tools Third Edition*. Germany: Springer.
- Kumar, V., Aksoy, L., Donkers, B., Venkatesan, R., Wiesel, T., & Tillmanns, S.** (2010). Undervalued or Overvalued Customers: Capturing Total Customer Engagement Value. *Journal of Service Research*, 13(3), 297-310.
- Lone, S., Harboul, N., & Weltevreden, J.** (2021). *2021 European E-commerce Report*. Hogeschool van Amsterdam.
- Madhoushi, Z., Hamdan, A. R., & Zainudin, S.** (2019, June). Aspect-Based Sentiment Analysis Methods in Recent Years. *Asia-Pacific Journal of Information Technology and Multimedia*, 7(2), 79-96.
- Masarifoglu, M., Tigrak, U., Hakyemez, S., Gul, G., Bozan, E., Buyuklu, A. H., & Özgür, A.** (2021, June). Sentiment Analysis of Customer Comments in Banking using BERT-based Approaches. *29th Signal Processing and Communications Applications Conference (SIU)*, 1-4.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J.** (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
- Morgan, N., & Rego, L. L.** (2006). The Value of Different Customer Satisfaction and Loyalty Metrics in Predicting Business Performance. *Marketing Science*, 25(5), 426-439.
- Özçift, A., Akarsu, K., Yumuk, F., & Söylemez, C.** (2021). Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): an empirical case study for Turkish. *Automatika*, 62(2), 226-238.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Duchesnay, E.** (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- Pennington, J., Socher, R., & Manning, C. D.** (2014). GloVe: Global Vectors for Word Representation. *In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532-1543.
- Piris, Y., & Gay, A.-C.** (2021). Customer satisfaction and natural language processing. *Journal of Business Research*, 124, 264-271.
- Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P.** (2016). SQuAD: 100,000+ Questions for Machine Comprehension of Text. *arXiv preprint arXiv:1606.05250*.
- Rambocas, M., & Pacheco, B. G.** (2018). Online sentiment analysis in marketing research: a review. *Journal of Research in Interactive Marketing*.

- Reichheld, F.** (2003). The One Number You Need to Grow. *Harvard business review*, 81(12), 46-55.
- Sarıççek, C., Yüksek, C., Ayhan, B., Yılmaz, S., Göktepe, F., & Erol, Ş.** (2022). *E-Pazaryerleri Platformları Sektör İncelemesi Nihai Raporu*. Ankara: Turkish Competition Authority.
- Schweter, S.** (2020, April). *BERTurk - bert models for Turkish*. Retrieved from <http://doi.org/10.5281/zenodo.3770924>
- Sığircı, İ. O., Özgür, H., Oluk, A., Uz, H., Çetiner, E., Oktay, H. U., & Erdemir, K.** (2020, September). Sentiment Analysis of Turkish Reviews on Google Play Store. *International Conference on Computer Science and Engineering (UBMK)*, 314-315. IEEE.
- Souza, F. D., & Souza Filho, J. d.** (2022). BERT for Sentiment Analysis: Pre-trained and Fine-Tuned Alternatives. *arXiv preprint arXiv:2201.03382*.
- Sun, C., Qiu, X., Xu, Y., & Huang, X.** (2019). How to Fine-Tune BERT for Text Classification? *China national conference on Chinese computational linguistics*, 194-206. Cham: Springer.
- Ucan, A., Naderalvojud, B., Sezer, E. A., & Sever, H.** (2016). SentiWordNet for New Language: Automatic Translation Approach. *12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 308-315. IEEE.
- United Nations.** (2021). Retrieved from COVID-19 and E-Commerce - A Global Review: https://unctad.org/system/files/official-document/dtlstict2020d13_en_0.pdf
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, I.** (2017). Attention Is All You Need. *Advances in neural information processing systems*, 30.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., . . . & Dean, J.** (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint*.
- Xiao, D., Ren, F., Pang, X., Cai, M., Wang, Q., He, M., . . . Fu, H.** (2021). A hierarchical and parallel framework for End-to-End Aspect-based Sentiment Analysis. *Neurocomputing*, 465, 549-560.
- Yang, C.-S., Chen, C.-H., & Chang, P.-C.** (2014). Harnessing consumer reviews for marketing intelligence: a domain-adapted sentiment classification approach. *Information Systems and e-Business Management*, 13(3), 403-419.

APPENDICES

APPENDIX A: Tables - Wallet Share Performance Metrics

APPENDIX B: Tables - Comparison of Actual and Predicted Wallet Shares



APPENDIX A

Table A.1 : Wallet Share Performance Metrics (2 Class – Type 1).

Category/Number of Customer	Case			Men's Perfume			Mobile Phone		
	100 Cus	300 Cus	500 Cus	100 Cus	300 Cus	500 Cus	100 Cus	300 Cus	500 Cus
MAE	0.0943	0.0314	0.0188	0.1484	0.0494	0.0296	0.0704	0.0234	0.0140
R2	0.9988	0.9988	0.9988	0.9998	0.9998	0.9998	0.9999	0.9999	0.9999
MAPE	0.0566	0.0566	0.0566	0.0197	0.0197	0.0197	0.0106	0.0106	0.0106
MSE	0.0261	0.0029	0.0010	0.0595	0.0066	0.0023	0.0166	0.0018	0.0006

Table A.2 : Wallet Share Performance Metrics (2 Class – Type 2).

Category/Number of Customer	Case			Men's Perfume			Mobile Phone		
	100 Cus	300 Cus	500 Cus	100 Cus	300 Cus	500 Cus	100 Cus	300 Cus	500 Cus
MAE	0.0758	0.0252	0.0151	0.0868	0.0289	0.0173	0.0392	0.0130	0.0078
R2	0.9994	0.9994	0.9994	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
MAPE	0.0427	0.0427	0.0427	0.0178	0.0178	0.0178	0.0079	0.0079	0.0079
MSE	0.0160	0.0017	0.0006	0.0179	0.0019	0.0007	0.0054	0.0006	0.0002

Table A.3 : Wallet Share Performance Metrics (5 Class).

Category/Number of Customer	Case			Men's Perfume			Mobile Phone		
	100 Cus	300 Cus	500 Cus	100 Cus	300 Cus	500 Cus	100 Cus	300 Cus	500 Cus
MAE	0.1336	0.0445	0.0267	0.3482	0.1160	0.0696	0.2020	0.0673	0.0404
R2	0.9982	0.9982	0.9982	0.9991	0.9991	0.9991	0.9994	0.9994	0.9994
MAPE	0.0976	0.0976	0.0976	0.0515	0.0515	0.0515	0.0540	0.0540	0.0540
MSE	0.0398	0.0044	0.0015	0.4204	0.0467	0.0168	0.0893	0.0099	0.0035

APPENDIX B**Table B.1 :** Comparison of Actual and Predicted Wallet Shares (2 Class – Type 1 & 100 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	376354	3.3727	3.3372	2107203	17.1836	16.9257	329366	19.4436	19.2551
2	759501	3.0708	3.2232	2018650	10.0665	10.2558	32315	8.6746	8.4813
3	1880900	3.0297	3.1578	2037808	8.4433	8.3103	920478	8.4599	8.3261
4	612418	2.7767	2.6326	1658884	5.2488	4.9623	42154	4.8543	4.9376
5	2343193	2.5484	2.5912	1747712	4.8082	4.7765	2809218	3.8428	3.7690
6	1696631	2.4758	2.4677	364365	4.7431	4.9380	2374001	3.2484	3.3384
7	104357	2.4447	2.5333	732566	4.3107	4.3158	728461	3.1055	3.0876
8	899464	2.2677	2.1837	3401700	4.1437	4.0647	246354	2.7404	2.7642
9	2411290	2.0290	2.1952	34142	3.6392	3.7515	572487	2.5666	2.8761
10	32310	1.9417	1.9628	801464	3.3035	3.1396	995645	2.4742	2.6908

Table B.2 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 1 & 300 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	1880900	6.2924	6.3671	2107203	28.3183	28.4008	329366	20.8333	20.7111
2	376354	5.5476	5.4352	2018650	10.2012	10.2968	920478	8.3722	8.3056
3	759501	4.9519	5.1211	2037808	7.7052	7.5726	32315	7.6444	7.6167
4	612418	3.0112	2.8934	1747712	5.2702	5.2155	42154	4.7667	4.8278
5	1696631	2.6134	2.5491	364365	4.5016	4.4722	728461	3.8611	3.8611
6	104357	2.5630	2.6065	1658884	3.8440	3.9472	2809218	3.0833	3.0111
7	2343193	2.4891	2.5034	3401700	3.7194	3.6000	515457	2.3278	2.3278
8	2411290	2.4490	2.5767	732566	3.1758	3.1702	2374001	2.3056	2.3333
9	899464	2.2940	2.2576	801464	3.0333	2.9992	246354	2.1889	2.2556
10	618452	1.9337	1.8278	34142	2.6968	2.6167	995645	1.6611	1.7167

Table B.3 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 1 & 500 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	1880900	7.7810	7.9590	2107203	26.1767	25.9800	329366	17.1362	17.1062
2	376354	7.3655	7.2557	2018650	10.5067	10.5733	32315	6.6033	6.5967
3	759501	5.9220	6.0511	2037808	6.6967	6.6800	920478	5.4933	5.5567
4	612418	3.4247	3.3262	1747712	4.0733	4.0300	42154	3.0667	3.0667
5	1696631	2.6910	2.6894	364365	3.6333	3.7400	728461	2.6500	2.6500
6	2343193	2.5796	2.6380	3401700	3.0400	2.9500	2809218	2.3333	2.3000
7	104357	2.4727	2.5463	1658884	2.7867	2.7500	2374001	2.1881	2.1881
8	2411290	2.2813	2.3635	732566	2.3600	2.3467	515457	1.8733	1.8567
9	618452	1.8696	1.8582	34142	2.2233	2.2067	572487	1.7833	1.8167
10	899464	1.8689	1.8154	801464	1.9200	1.8700	995645	1.6233	1.6067

Table B.4 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 2 & 100 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	759501	3.9340	4.0785	2107203	21.0679	20.1985	329366	20.4353	19.5171
2	1880900	3.7665	3.8414	2018650	12.9469	12.9124	920478	8.9278	9.0032
3	376354	3.6721	3.7758	2037808	8.7782	8.6223	32315	7.6873	7.7984
4	612418	3.3288	3.3667	1747712	5.6996	5.5560	42154	4.3647	4.6925
5	1696631	2.8113	2.7957	364365	5.2715	5.6545	728461	3.3242	3.2909
6	899464	2.6197	2.5629	1658884	4.4322	4.2953	2809218	3.2262	3.1929
7	104357	2.4580	2.4427	3401700	3.5433	3.3448	995645	3.1099	3.0028
8	2343193	2.2950	2.3653	801464	3.5048	3.8202	515457	2.9413	2.8774
9	2411290	2.2583	2.3402	732566	3.2568	3.3410	572487	2.7964	2.8476
10	32310	1.9609	1.9627	34142	2.9230	2.9996	2374001	2.7218	3.0948

Table B.5 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 2 & 300 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	1880900	6.9234	7.0154	2107203	28.3270	28.0373	329366	23.4111	23.2278
2	759501	5.9369	6.0049	2018650	11.3175	11.5397	920478	7.1167	7.0611
3	376354	5.7553	5.8075	2037808	8.1373	8.2714	32315	6.5389	6.5778
4	612418	3.4058	3.4015	1747712	4.2913	4.1056	42154	3.3056	3.3333
5	1696631	3.1374	3.1632	3401700	4.1556	4.0444	2809218	3.2500	3.3611
6	2411290	2.7133	2.7739	364365	3.5468	3.5968	728461	3.0278	3.0278
7	899464	2.6950	2.6567	732566	3.0611	3.1857	2374001	2.8056	2.8056
8	104357	2.5420	2.5285	1658884	3.0333	3.0500	515457	2.5556	2.5000
9	2343193	2.3032	2.3394	801464	2.9000	2.9833	995645	2.5500	2.5500
10	618452	2.1192	2.1186	34142	2.4556	2.4111	1993058	1.6222	1.5944

Table B.6 : Comparison of Actual and Predicted Wallet Shares (2 Class – Type 2 & 500 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	1880900	9.3120	9.3609	2107203	25.9714	25.8033	329366	17.7067	17.5733
2	376354	6.8349	6.9194	2018650	9.7514	9.8733	920478	5.5833	5.5833
3	759501	6.2943	6.4172	2037808	6.3981	6.2733	32315	5.2400	5.2400
4	1696631	3.0983	3.0792	1747712	3.9233	3.8833	42154	3.4667	3.5000
5	612418	2.9342	2.9647	364365	3.2900	3.3167	728461	2.2500	2.2500
6	899464	2.6035	2.5473	3401700	2.6200	2.6133	2809218	2.1333	2.1333
7	104357	2.3589	2.3442	1658884	2.5967	2.5533	2374001	2.0333	2.0667
8	2411290	2.3525	2.3935	732566	2.5733	2.6000	515457	1.8667	1.8667
9	2343193	2.3060	2.3392	801464	2.0600	2.1367	995645	1.7333	1.7333
10	618452	1.9434	1.9414	2628904	2.0433	2.0433	1672829	1.4667	1.4667

Table B.7 : Comparison of Actual and Predicted Wallet Shares (3 Class & 100 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	376354	3.1704	3.2238	2107203	17.0181	16.0620	329366	19.2309	18.8027
2	612418	2.9964	3.0785	2018650	11.4558	11.6152	32315	8.9439	8.5678
3	1880900	2.7936	2.7879	2037808	7.4138	7.5150	920478	7.5419	7.4754
4	759501	2.7584	2.8151	1747712	6.3488	6.5613	42154	4.8785	4.8777
5	1696631	2.5674	2.3811	364365	5.5655	5.8131	728461	3.6117	3.5400
6	2343193	2.3336	2.4216	1658884	4.7411	4.7089	2809218	3.5663	3.4778
7	104357	2.3101	2.4600	3401700	4.6510	4.5305	995645	2.8028	2.7956
8	899464	2.2359	2.0804	732566	3.8034	3.9058	515457	2.6459	2.5686
9	2411290	2.0038	2.0655	431387	3.4261	3.5543	2374001	2.4694	2.5512
10	618452	1.8067	1.8769	34142	3.0927	3.4133	572487	2.2210	2.4948

Table B.8 : Comparison of Actual and Predicted Wallet Shares (3 Class & 300 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	1880900	6.5819	6.5733	2107203	26.1224	25.7694	329366	20.4135	20.0722
2	376354	5.9216	5.9477	2018650	10.9438	11.0925	32315	7.6556	7.5667
3	759501	5.1713	5.3510	2037808	6.6882	6.7131	920478	6.0413	5.9333
4	612418	3.1062	3.1485	1747712	4.9167	5.1111	42154	3.7667	3.8556
5	1696631	2.8538	2.6697	364365	4.6224	4.7353	728461	2.6659	2.7222
6	2343193	2.6295	2.6536	1658884	3.5024	3.4659	515457	2.5468	2.5611
7	104357	2.4121	2.6049	3401700	3.2312	3.2488	246354	2.4889	2.5222
8	2411290	2.3562	2.4105	801464	2.8579	2.8524	2374001	2.1413	2.1056
9	899464	1.9664	1.8548	732566	2.7254	2.7040	2809218	2.0968	2.1000
10	1578605	1.8398	1.7626	34142	2.3485	2.5060	572487	1.9500	2.0000

Table B.9 : Comparison of Actual and Predicted Wallet Shares (5 Class & 100 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	376354	3.1505	3.1664	2107203	17.0781	15.3926	329366	18.5837	18.2091
2	759501	2.9768	3.0838	2018650	10.5258	10.8889	32315	8.7178	8.8893
3	612418	2.9239	2.7581	2037808	7.4949	6.9398	920478	7.2448	7.3143
4	1880900	2.8567	2.7932	1747712	6.0632	6.2557	42154	5.1297	5.1361
5	1696631	2.6477	2.5950	1658884	5.0517	4.8826	2809218	3.9404	3.6849
6	899464	2.5415	2.3064	364365	5.0258	5.4767	728461	3.9147	3.9845
7	104357	2.1971	2.4776	3401700	4.4434	4.1616	2374001	3.1428	3.2659
8	2343193	2.0158	1.9881	34142	3.6878	4.4310	515457	2.6579	2.8821
9	2411290	1.9339	1.8824	431387	3.3172	3.3633	995645	2.5976	2.5556
10	618452	1.7165	1.7333	732566	3.2338	3.6813	228373	2.5560	2.1139

Table B.10 : Comparison of Actual and Predicted Wallet Shares (5 Class & 300 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	1880900	6.4912	6.5234	2107203	25.8627	25.0254	329366	20.8754	20.4048
2	376354	5.9053	6.1386	2018650	12.0063	11.7865	32315	7.9921	8.3079
3	759501	5.0599	5.2581	2037808	7.7071	7.5032	920478	6.9270	6.7833
4	612418	3.6647	3.4853	1747712	5.5730	5.7651	42154	4.1833	4.3833
5	104357	2.7163	2.8070	364365	4.2587	4.3540	728461	3.1000	3.0833
6	1696631	2.6333	2.5434	3401700	4.1579	4.0873	515457	3.0603	3.0802
7	2411290	2.5093	2.4820	1658884	4.0413	4.0048	2809218	2.6706	2.5127
8	899464	2.2350	2.0307	732566	2.9484	3.1619	2374001	2.3325	2.3302
9	2343193	2.0657	2.1090	801464	2.7056	2.8500	995645	2.0056	1.8111
10	618452	1.8267	1.8759	34142	2.4317	2.7468	246354	1.9817	2.0857

Table B.11 : Comparison of Actual and Predicted Wallet Shares (5 Class & 500 Cus).

Order	Case			Men's Perfume			Mobile Phone		
	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)	Seller_Id	Actual Wallet Share (%)	Predicted Wallet Share (%)
1	1880900	7.7980	7.8893	2107203	25.3745	24.8912	329366	17.7567	17.3367
2	376354	6.6761	6.9797	2018650	11.4055	11.4421	32315	6.1233	6.1133
3	759501	5.7316	5.7449	2037808	6.6221	6.6455	920478	5.8300	5.7633
4	612418	3.3886	3.2460	1747712	4.5726	4.6860	42154	3.5000	3.5667
5	1696631	2.6619	2.6070	364365	3.4967	3.3700	728461	2.5367	2.5500
6	104357	2.5665	2.7283	1658884	3.3433	3.3467	2809218	2.0733	2.0667
7	2411290	2.3021	2.2576	3401700	3.1855	3.0200	515457	2.0167	2.0833
8	2343193	2.2515	2.2742	801464	2.7633	2.8467	2374001	2.0067	1.9800
9	899464	2.0317	1.8824	732566	2.4133	2.5633	1993058	1.5667	1.5667
10	618452	1.7516	1.7874	431387	1.8505	1.8426	572487	1.5467	1.6100



CURRICULUM VITAE

Name Surname : Muhammet Ali KADIOĞLU

EDUCATION

- **B.Sc.:** 2019, Yıldız Technical University, Faculty of Mechanical Engineering, Industrial Engineering Department
- **M.Sc.:** 2022, Istanbul Technical University, Graduate School of Science Engineering and Technology, Industrial Engineering Department