ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

FUZZY CLUSTERING BASED ENSEMBLE LEARNING APPROACH: APPLICATIONS IN DIGITAL ADVERTISING

Ph.D. THESIS

Ahmet Tezcan TEKİN

Department of Management Engineering

Management Engineering Programme

DECEMBER 2021

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

FUZZY CLUSTERING BASED ENSEMBLE LEARNING APPROACH: APPLICATIONS IN DIGITAL ADVERTISING

Ph.D. THESIS

Ahmet Tezcan TEKİN
(507172006)

Department of Management Engineering

Management Engineering Programme

Thesis Advisor: Prof. Dr. Ferhan ÇEBİ
Thesis Co-Advisor: Prof. Dr. Tolga KAYA

DECEMBER 2021

ISTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

BULANIK KÜMELEME TABANLI TOPLULUK ÖĞRENMESİ YAKLAŞIMI: DİJİTAL REKLAM ALANINDA UYGULAMALAR

DOKTORA TEZİ

Ahmet Tezcan TEKİN
(507172006)

İşletme Mühendisliği Anabilim Dalı

İşletme Mühendisliği Programı

Tez Danışmanı: Prof. Dr. Ferhan ÇEBİ
Eş Danışman: Prof. Dr. Tolga KAYA

ARALIK 2021

Ahmet Tezcan Tekin, a Ph.D. student of ITU Graduate School student ID 507172006 successfully defended the thesis/dissertation entitled "FUZZY CLUSTERING BASED ENSEMBLE LEARNING APPROACH: APPLICATIONS IN DIGITAL ADVERTISING", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Ferhan ÇEBİ** Istanbul Technical University	.....
Co-advisor :	**Prof. Dr. Tolga KAYA** Istanbul Technical University	.....
Jury Members :	**Prof. Dr. Mehmet Mutlu YENİSEY** Istanbul University	.....
	Prof. Dr. Dilay Çelebi GONIDIS Istanbul Technical University	.....
	Prof. Dr. Başar ÖZTAYŞİ Istanbul Technical University	.....
	Asst. Prof. Dr. Cemal Okan ŞAKAR Bahçeşehir University	.....
	Asst. Prof. Dr. Barış SOYBİLGEN Bilgi University	.....

Date of Submission : 16 November 2021
Date of Defense : 27 December 2021

To my family,

FOREWORD

During my Ph.D thesis, many people have contributed to this period in different ways. First of all, I would like to thank my advisor Prof. Dr. Ferhan ÇEBİ and my co-advisor Prof. Dr. Tolga KAYA for their valuable help and support in this period. I would also thank them for sharing their knowledge and experience in the academic field.

I also want to express my sincere thanks to Prof. Dr. Başar ÖZTAYŞİ for his support and his guidance. Additionally, I would like to thank Asst. Prof. Dr. Cemal Okan ŞAKAR and Asst. Prof. Dr. Barış SOYBİLGEN, for their valuable comments and suggestions on the draft of this thesis.

I am also thankful to my family for their motivation during this challenging Ph.D. process.

December 2021

Ahmet Tezcan TEKİN
(M.Sc. Management Engineer)

TABLE OF CONTENTS

Page

FOREWORD ..... **ix**
TABLE OF CONTENTS ..... **xi**
ABBREVIATIONS ..... **xiii**
SYMBOLS ..... **xv**
LIST OF TABLES ..... **xvii**
SUMMARY ..... **xix**
ÖZET ..... **xxiii**
1. INTRODUCTION **1**
1.1 Purpose of Thesis 3
2. CLICK AND SALES PREDICTION FOR OTAS' DIGITAL ADVERTISEMENTS: FUZZY CLUSTERING BASED APPROACH **7**
2.1 Introduction 7
2.2 Modelling 9
2.3 Proposed Methodology 12
2.3.1 Regression based approach 12
2.3.2 Clustering based and regression based approach 18
2.4 Conclusion 20
3. CUSTOMER LIFETIME VALUE PREDICTION FOR GAMING INDUSTRY: FUZZY CLUSTERING BASED APPROACH **23**
3.1 Introduction 23
3.2 Literature Review 24
3.3 Proposed Methodology and Modelling 31
3.4 Conclusion 37
4. RETENTION PREDICTION IN THE GAMING INDUSTRY: FUZZY MACHINE LEARNING APPROACH **39**
4.1 Literature Review 41
4.1.1 Ensemble learning 41
4.1.1.1 Bagging 41
4.1.1.2 Boosting 42
4.1.2 Evaluation of performance metrics 44
4.1.3 Fuzzy clustering techniques 45
4.1.4 Proposed methodology 47
4.1.5 Conclusion 50
5. CONCLUSIONS **53**
REFERENCES **57**
CURRICULUM VITAE **65**

ABBREVIATIONS

Adaboost : Adaptive Boosting
CART : Classification and Regression Trees
CB : Catboost
CTR : Click-Through Rate
D : Depth
F1 Score : F-Score
FCM : Fuzzy C-Means Clustering
FPC : Fuzzy Partition Coefficient
FPCM : Fuzzy Possibilistic C-Means Clustering
GBC : Gradient Boosting Classifier
GBDT : Gradient Boosting Decision Trees
LGBM : Light Gradient Boosting Machine
LightGBM : Light Gradient Boosting Machine
LR : Learning Rate
MAE : Mean Absolute Error
MD : Maximum Depth
NE : Number of Estimators
OTA : Online Travel Agency
PCM : Possibilistic C-Means Clustering
PFCM : Possibilistic Fuzzy C-Means Clustering
PPC : Pay-Per Click
$\boldsymbol{R^2}$: R-Squared Value
RF : Random Forest
SVR : Support Vector Regressor
XGB : Extreme Gradient Boosting
XGBoost : Extreme Gradient Boosting

SYMBOLS

$\boldsymbol{D}$	**:** Dataset
$\boldsymbol{n}$	**:** Number of Samples
$\boldsymbol{C^*}$	**:** Classifier
$\boldsymbol{f_t(x_i)}$	**:** Output Function of Explanatory Variable $\boldsymbol{x_i}$
$\boldsymbol{R_j}$	**:** Disjoint Region
$\boldsymbol{W_{(x_i)}}$	**:** Weight of the Prediction of $\boldsymbol{x_i}$
$\boldsymbol{k}$	**:** Initial Seed Value of Cluster
$\boldsymbol{\hat{m}_j}$	**:** Individual Tree
$\boldsymbol{E}$	**:** Error Function
$\boldsymbol{F_{t-1}(x)}$	**:** Previous Learner
$\boldsymbol{\beta_j}$	**:** Expasion Coefficient
$\boldsymbol{\mu}$	**:** Degree of Membership
$\boldsymbol{acc}$	**:** Accuracy

LIST OF TABLES

Page

Table 2.1: Base features used in CTR and impression prediction. 14
Table 2.1 (continued): Base features used in CTR and impression prediction. 15
Table 2.2: Algortihm's results for CTR prediction. 15
Table 2.3: Algortihm's results for impression prediction. 16
Table 2.4: Algortihm's results for click prediction by predicted impression multiplied by predicted CTR. 16
Table 2.5: The additional features which are used in the sales prediction. 18
Table 2.6: Algorithms' results for sales prediction. 18
Table 2.7: Number of set for clusters. 19
Table 2.8: Comparison of clustering algorithms for CTR prediction. 19
Table 2.9: Clustering algorithms and regression algorithms' results for impression prediction. 20
Table 2.10: Clustering algorithms and regression algorithms' results for click prediction by predicted impression multiplied by predicted CTR. 20
Table 3.1: Base features used in lifetime value prediction. 31
Table 3.1 (continued): Base features used in lifetime value prediction. 32
Table 3.2: Additional features used in lifetime value prediction. 33
Table 3.3: Fuzzy cluster details. 35
Table 3.4: Model and parameters prediction results on groups (RMSE). 36
Table 3.5: Overall prediction results with each model and parameters (RMSE). 36
Table 4.1: Algorithm performances with default parameters. 47
Table 4.2: Base features used in retention prediction. 48
Table 4.3: Fuzzy cluster details. 49
Table 4.4: Algorithm performances with different parameters. 49
Table 4.5: Overall prediction results with each model and parameters (accuracy). 50

FUZZY CLUSTERING BASED ENSEMBLE LEARNING APPROACH: APPLICATIONS IN DIGITAL ADVERTISING

SUMMARY

Although the history of machine learning algorithms is quite old, it has been popularly used in the last ten years. The main reason for this situation is that it has become possible to run these algorithms even on our personal computers with the developing computer hardware technology. In addition, the size of the data generated in the internet environment is increasing exponentially with each passing day, with digitalization and internet usage becoming more widespread. Therefore, the need for technologies such as big data and machine learning is increasing day by day. In line with the increasing demands, machine learning has become an indispensable need in academia and the private sector. Thanks to machine learning, companies make predictions about their future processes, thus aiming to eliminate future uncertain situations and create more effective process management. E.g., A company seeks to use its marketing budget more effectively by using machine learning technologies for its marketing processes and thus maximizing its profitability rate.

In recent years, there have been many studies in the literature on the development of machine learning algorithms and the elimination of the weaknesses of traditional machine learning methods. Regardless of the type of problem in the prediction process, the aim is to predict with a minimum error rate. In this context, many methods have been tried. The ensemble learning approach is one of the most successful methods in the literature, proving its success for this purpose. The purpose of ensemble learning is to combine multiple algorithms to close each other's weaknesses and increase the success rate in prediction.

Observations on the dataset to be estimated may be characteristically similar or very different from each other. In this case, in many studies in the literature, the clustering process is performed before applying machine learning algorithms, and then the modelling stage is started. In such approaches, hard clustering approaches are used. Hard clustering approaches assign each observation value to only one cluster due to their working principles. Therefore, the sizes of the subsets to be modelled in some cases do not reach the size of the training set required for higher prediction success to occur. Considering that an observation value contains the characteristics of more than one cluster simultaneously, it is seen that the soft clustering approach is used to eliminate this problem. Although there are many studies in the literature on the fuzzy clustering method, which is a part of the soft clustering approach, there are not many examples in the literature regarding the use of the machine learning approach as an intermediate method in terms of improving its results. In this thesis, after the fuzzy clustering approach applied to the observation set with three published essays, it is aimed to ensemble the most successful models of each cluster, taking into account the error rates and thus improving the model performances. To test the validity of this approach, different studies were carried out for both regression and classification problems with datasets obtained from different sectors.

In the first study, click and sales predictions were realised using digital advertisement performance data and reservation data in metasearch engines of an online travel agency operating in Turkey. This prediction is crucial for the company's short, medium and long-term financial goals. In this study, the traditional regression method and the proposed fuzzy clustering approach were used together and the results were compared with the results of the traditional methods. Machine learning algorithms were applied directly to the dataset, which had been applied data preprocessing and feature engineering within the framework of traditional methods. Then the modelling study was carried out again after the hard clustering and soft clustering approaches were applied to the dataset. As a result, although the processing load increased due to the inclusion of the clustering approach in addition to the modelling stage, more effective results were obtained than applying machine learning algorithms directly to the dataset. At the same time, the results obtained after the hard clustering approach and fuzzy clustering approaches were compared. It was observed that the success rate of the predictions made after the fuzzy clustering approach was higher.

In the second study, the approach proposed in the first article is tested for solving a different problem with different sector data. In this study, it has been tried to predict the lifetime value of the customers by using the game data and session information of the users of a mobile crossword puzzle game published in more than thirty languages and more than thirty countries. Ensemble learning algorithms, which were not used in the first article, were examined in more depth and focused on algorithms that could achieve higher prediction success rates when used together with fuzzy clustering. Different hyperparameter combinations of Catboost, Extreme Gradient Boosting and Light Gradient Boosting algorithms, which are seen in the literature to be generally more successful than traditional machine learning algorithms, were tested separately for each cluster after the clustering phase performed with the Fuzzy C-Means clustering algorithm. The prediction values of the three most successful of these combinations were weighted to be inversely proportional to the error rates, and the error rates of the resulting predictions were compared with the results of other model-parameter combinations. It has been determined that the model established with the proposed method has a lower error rate than other models, thus making a more efficient prediction.

In the third study, customer retention rate prediction was carried out with a different dataset collected in the gaming industry. Unlike the first two studies, in this study, a classification problem was tried to be solved with the proposed method, at the same time, different cluster initial parameters and different fuzziness parameters were tested. The aim is to obtain a more optimal clustering in the Fuzzy C-Means clustering approach, and the clustering process was the most successful combination. Since the nature of the problem is a classification problem, the prediction was carried out by weighting the accuracy results instead of the error rates of the algorithms at the stage of combining the results of the algorithm-parameter combinations. As a result of this study, it has been observed that the results of the method applied on different clusters clustered with the fuzzy clustering approach produce more effective results than applying machine learning algorithms directly to the dataset.

As a result, this thesis provides the opportunity to make more successful predictions in datasets with different characteristics by strengthening the concept of ensemble learning, which has an important place in developing machine learning approaches with fuzzy clustering approaches. In addition, it allows identifying observation sets

that contain the characteristics of more than one cluster simultaneously and to model in separate clusters during the modelling phase to create more effective prediction results.

In this constantly developing field, new studies can progress from many branches. First of all, in the fuzzy clustering stage, instead of the Fuzzy C-Means clustering method, other alternative fuzzy clustering approaches in the literature can be tried again during the modelling stage. And a different fuzzy clustering algorithm can be preferred according to the efficiency result. At the same time, it may be possible to change the weight coefficients with different methods or shapes at the stage of combining the results of the most successful models. Beyond all this, this method will enable to produce more effective results by using it together with new machine learning algorithms that will be introduced to the literature in the future.

BULANIK KÜMELEME TABANLI TOPLULUK ÖĞRENMESİ YAKLAŞIMI: DİJİTAL REKLAM ALANINDA UYGULAMALAR

ÖZET

Makine öğrenmesi algoritmalarının geçmişi oldukça eskiye dayansa da popüler olarak kullanıma son on yılda başlanmıştır. Bu durumun temel nedeni gelişen bilgisayar donanım teknolojisi ile kişisel bilgisayarlarımızda dahi bu algoritmaları çalıştırmak mümkün hale gelmiştir. Ayrıca, dijitalleşme ve internet kullanımının her geçen gün daha yaygın hale gelmesi ile internet ortamında oluşan verinin boyutu her geçen gün katlanarak artmaktadır. Bu nedenle büyük veri ve makine öğrenmesi gibi teknolojilere ihtiyaç her geçen gün artmaktadır. Artan ihtiyaçlar doğrultusunda makine öğrenmesi sadece akademik alanda değil özel sektör içinde de vazgeçilmez bir ihtiyaca dönüşmüştür. Makine öğrenmesi sayesinde şirketler gelecekteki süreçleri ile ilgili tahminlemeler yapmakta, bu sayede gelecekteki belirsiz durumları ortadan kaldırıp daha etkin bir süreç yönetimi yapmayı amaçlamaktadırlar. Örneğin; bir şirket pazarlama süreçleri için makine öğrenmesi teknolojilerini kullanarak elindeki pazarlama bütçesini daha etkin kullanmayı amaçlamakta ve bu sayede karlılık oranını maksimize etmeye çalışmaktadır.

Son yıllarda makine öğrenmesi algoritmalarının geliştirilmesi, geleneksel makine öğrenmesi yöntemlerinin zayıf yönlerinin giderilmesi noktasında literatürde birçok çalışma bulunmaktadır. Tahminleme işleminde problemin türü ne olursa olsun amaç minimum hata oranı ile tahminleme yapmaktır. Bu kapsamda birçok yöntem denenmiştir. Topluluk öğrenmesi yaklaşımı literatür bu amaç için başarısını kanıtlayan en başarılı yöntemlerden biridir. Topluluk öğrenmesinin amacı, birden çok algoritmayı birleştirerek birbirlerinin zayıf yönlerini kapatmak ve tahminlemedeki başarı oranını artırmaktır.

Tahminleme yapılacak veriseti üzerindeki gözlemler karakteristik olarak birbilerine benzerlik gösterebilir ya da birbirlerinden çok farklı olabilirler. Bu durumda litetaürde birçok çalışmada makine öğrenmesi algoritmaları uygulanmadan önce kümeleme işlemi gerçekleştirilip daha sonra modelleme aşamasına geçilmektedir. Bu tür yaklaşımlarda daha çok katı kümeleme yaklaşımları kullanılmaktadır. Katı kümeleme yaklaşımları, çalışma prensipleri gereği her bir gözlem değerini yalnızca bir kümeye atamaktadırlar. Dolayısıyla modelleme yapılacak alt kümelerin boyutları bazı durumlarda yüksek başarılı tahminlemenin gerçekleşmesi için gereken öğrenme seti boyutuna ulaşamamaktadır. Bir gözlem değerinin aynı anda birden fazla kümeye ait karakteristik özellikleri barındırma durumu göz önünde bulundurulduğunda esnek kümeleme yaklaşımının bu problemi ortadan kaldırmak için kullanıldığı görülmektedir. Esnek kümeleme yaklaşımının bir parçası olan bulanık kümeleme yöntemine ait literatürde birçok çalışma olsa da makine öğrenmesi yaklaşımının sonuçlarını iyileştirme anlamında ara yöntem olarak kullanılmasına ilişkin literatürde çok fazla örnek bulunmamaktadır. Bu tezde, yayınlanmış üç makale ile gözlem setine uygulanan bulanık kümeleme yaklaşımı sonrasında her bir kümeye ait en başarılı modellerin hata oranları göz önünde bulundurularak topluluk haline getirilmesi ve bu

sayede model performanslarının iyileştirilmesi amaçlanmıştır. Uygulanan bu yaklaşımın geçerliliğinin test edilmesi amacı ile farklı sektörlerden elde edilen verisetleri ile birlikte hem regresyon hem de sınıflandırma problemleri için farklı çalışmalar gerçekleştirilmiştir.

İlk makalede Türkiye'de faaliyet gösteren bir çevrimiçi seyahat acentasının meta arama motorlarındaki dijital reklam performans verileri ile rezervasyon verileri kullanılarak tıklama ve satış tahminlemesi gerçekleştirilmeye çalışılmıştır. Bu tahminleme, firmanın kısa, orta ve uzun vadedeli finansal hedefleri için oldukça önem arz etmektedir. Bu çalışmada geleneksel regresyon yöntemi ile önerilen bulanık kümeleme yaklaşımı birlikte kullanılmış ve sonuçlar geleneksel yöntem sonuçları ile karşılaştırılmıştır. Geleneksel yöntemler çerçevesinde veri ön işleme ve öznitelik mühendisliği uygulanmış olan verisetine direkt olarak makine öğrenmesi algoritmaları uygulanmış, sonrasında ise katı kümeleme ve esnek kümeleme yaklaşımları verisetine uygulandıktan sonra modelleme çalışması tekrardan gerçekleştirilmiştir. Sonuç olarak modelleme aşamasına ek olarak kümeleme yaklaşımının da dahil edilmesi sonucunda işlem yük artmış olsa da verisetine direkt olarak makine öğrenmesi algoritmalarını uygulamaya göre daha etkin sonuçlar elde edilmiştir. Aynı zamanda katı kümeleme yaklaşımı ile bulanık kümeleme yaklaşımları sonrasında elde edilen sonuçlar da karşılaştırılmış ve bulanık kümeleme yaklaşımı sonrasında yapılan tahminlerin başarı oranının daha yüksek olduğu gözlemlenmiştir.

İkinci makalede, ilk makalede önerilen yaklaşım farklı bir sektör verisi ile farklı bir problemin çözümü için test edilmiştir. Bu çalışmada otuzdan fazla dil seçeneği ve otuzdan fazla ülkede yayında olan bir mobil çapraz bulmaca oyununun kullanıcılarının ilk yirmidört saatteki oyun verileri ve oturum bilgileri kullanılarak müşterilerin yaşam ömrü tahmin edilmeye çalışılmıştır. İlk makalede kullanılmayan topluluk öğrenmesi algoritmaları daha derinlemesine irdelenmiş, bulanık kümeleme ile birlikte kullanılması sonucunda daha yüksek tahminleme başarı oranına ulaşabilecek algoritmalar üzerine yoğunlaşılmıştır. Bulanık C-Ortalamalar kümeleme algoritması ile gerçekleştirilen kümeleme aşamasından sonra geleneksel makine öğrenmesi algoritmalarına göre genelde daha başarılı oldukları literatürde de görülen Catboost, Extreme Gradient Boosting ve Light Gradient Boosting algoritmalarının farklı hiperparametre kombinasyonları her bir küme için ayrı ayrı denenmiştir. Bu kombinasyonların en başarılı üç tanesinin tahmin değerleri hata oranları ile ters orantılı olacak şekilde ağırlandırılarak birleştirilmiş ve sonuç tahminlerin hata oranları diğer model-parametre kombinasyonlarının sonuçları ile karşılaştırılmıştır. Önerilen yöntemle kurulmuş olan modelin diğer modellere göre daha düşük hata oranına sahip olduğu dolayısıyla daha etkin tahminleme yaptığı tespit edilmiştir.

Üçüncü makalede ise yine oyun sektöründe toplanan farklı bir veriseti ile müşteri muhafaza oranı tahminlemesi gerçekleştirilmiştir. İlk iki çalışmadan farklı olarak, bu çalışmada bir sınıflandırma problemi önerilen yöntem ile çözülmeye çalışılmış, aynı zamanda Bulanık C-Ortalamalar kümeleme yaklaşımında daha optimum bir kümeleme elde edilmesi ama amacı ile farklı küme başlangıç parametreleri ile farklı bulanıklık parametreleri test edilmiş ve kümeleme işlemi en başarılı kombinasyon ile gerçekleştirilmiştir. Problemin niteliği bir sınıflandırma problemi olduğu için algoritma-parametre kombinasyonlarının sonuçlarının birleştirilmesi aşamasında algoritmaların hata oranları yerine doğruluk sonuçları ağırlıklandırılarak tahminleme gerçekleştirilmiştir. Bu çalışmanın sonucunda bulanık kümeleme yaklaşımı ile kümelenmiş farklı kümeler üzerine uygulanan yöntemin sonuçlarının makine

öğrenmesi algoritmalarını direkt olarak verisetine uygulamaya göre daha etkin sonuçlar ürettiği gözlemlenmiştir.

Sonuç olarak, bu tez gelişmekte olan makine öğrenmesi yaklaşımlarında önemli bir yere sahip olan topluluk öğrenmesi kavramının bulanık kümeleme yaklaşımı ile güçlendirilerek farklı karakteristik özelliklere sahip olan veri kümelerinde daha başarılı bir tahminleme yapma olanağı sunmaktadır. Buna ek olarak, aynı anda birden fazla kümenin karakteristik özelliklerini barındıran gözlem setlerinin tespit edilmesi ve modelleme aşamasında ayrı kümelerde modellenip daha etkin tahmin sonuçlarının oluşturulmasına imkan tanımaktadır.

Sürekli olarak gelişmekte olan bu alanda yeni çalışmalar pek çok daldan ilerleyebilir. Öncelikle bulanık kümeleme aşamasında Bulanık C-Ortalamalar kümeleme yöntemi yerine literatürdeki diğer alternatif bulanık kümeleme yaklaşımları yine modelleme aşamasında denenip, etkinlik sonucuna göre farklı bulanık kümeleme algoritması tercih edilebilir. Aynı zamanda en başarılı modellerin sonuçlarının birleştirilmesi aşamasında farklı yöntemler veya şekiller ile ağırlık katsayılarının değiştirilmesi mümkün olabilir. Tüm bunların ötesinde bu yöntem ilerleyen zamanlarda literatüre kazandırılacak olan yeni makine öğrenmesi algoritmaları ile birlikte kullanılarak daha etkin sonuçlar üretilmesine imkan tanıyacaktır.

1. INTRODUCTION

When the history of machine learning algorithms is examined, although their introduction to the literature is quite old, their frequency of use has increased, especially with the development of technology and the increase in the number of operations that computers can perform per second. Both academic and private-sector problems are trying to be solved with different machine learning methods and the popularity of the subject is increasing day by day. The size and content of the data, especially in the digital environment, is growing rapidly. If these data can be processed, the results can shed light on the solution to many problems.

Machine learning problems, which are generally tried to be solved in the private sector, are demand forecasting, sales forecasting, cost forecasting, customer lifetime value, etc. As a result, whatever the forecast is, the aim is to support the decision-making process and predict many uncertain events in the future. This prediction process has an essential place in shaping the information and strategic operations of the companies for the situations they will encounter in the future. Regardless of the type of problem in the estimation process, the aim is to have a minimum error rate in the prediction.

Many methods have been tried in the literature to achieve this purpose and one of the most applied methods today is the ensemble learning method. The fundamental idea behind ensemble learning is to combine a series of models, each of which solves the same original problem, to produce a superior composite global model with more precise and dependable predictions than a single model can provide [1]. Ensemble learning algorithms have a wide range of usage in the literature. Matloob et al. [81] used ensemble learning methods in software defect prediction. Weeraddana et al. [82] used based novel ensemble learning framework for electricity operational forecasting. Also, there are many studies in the literature that use ensemble learning technics to improve the prediction accuracy in the field of biology to detect counterfeit banknotes[83–90]. Matloob et al. [81] state that ensemble learning techniques generally perform significantly better than individual algorithms.

Clustering algorithms can also benefit from ensemble approaches to improve their quality and robustness [2]. This method aims to close the weaknesses of multiple algorithms by combining them and increasing the success rate in prediction. For this purpose, regardless of whether the problem is classification or regression, there are multiple methods proposed and proven successful in the literature.

Unsupervised classification, also known as clustering, identifies classes from data without knowing the labels. Clustering algorithms are used to find groupings of items, known as clusters, that are more similar to one another than to other clusters. This method of data analysis is closely connected to the process of developing a data model, which entails establishing a reduced set of attributes that may be used to offer an intuitive explanation of significant parts of a dataset [3]. The clustering approach, a sub-branch of machine learning, can be combined with a classification or regression problem. At the same time, it serves to cluster the values in the observation set on its own. The purpose here is; Again, it is to increase the success rate by applying machine learning algorithms on observation sets that are similar to each other. Therefore, many studies in the literature combine unsupervised and supervised learning. Some researchers have proposed semi-supervised learning methods that attempt to improve the performance of supervised learning models by utilizing the information included in unlabeled data [91-93]. On the other hand, Cui et al. [94] proposed a novel clustering-based intelligent ensemble learning method that analyses the data's internal relationships for focusing on samples belonging to different categories in the same cluster. They aim to improve the generalization ability of the classification system.

The concept of fuzzy logic proposed by Zadeh [4] is also a widespread method for machine learning problems. One of the main reasons for this is that the answers to real-life problems do not consist of only 0 and 1 singular values, as in machine learning approaches that deal with real-life problems. The concept of fuzzy logic is used as fuzzy machine learning machine learning problems. So, there are many studies in the literature with the fuzzy extension of standard, non-fuzzy methods: from rule induction to fuzzy rule induction [95]. With the fuzzy extension, decision trees as fuzzy decision trees [96], nearest neighbour classifier as fuzzy nearest neighbour classifier [97], support vector machine as fuzzy support vector machine

[98] are used in the literature. We are trying to explain the answers to many problems in current life with values between 0 and 1 with the power of fuzzy logic.

In this study, the estimation problem is mainly addressed to a specific group of products, people, etc. Fuzzy logic was used for bagging and boosting, one of the ensemble learning methods in the literature, in machine learning problems with a set of observations that can be clustered. In detail, the training set was clustered. The prediction was made with different parameters of more than one ensemble learning algorithm and the optimum result was tried to be reached with these ensemble learning members. This method aims to determine if the items that are product, person, etc. which are similar to each other in terms of behavioural characteristics, belong to more than one cluster at the same time by subjecting them to fuzzy clustering. Also, this method aims to make predictions by making the algorithms with the most optimum validation rate for these cluster variations being members of the ensemble by the percentage of success rates. In this way, the prediction process will be made for items similar to each other via considering the cluster success, and the overall success of the model has been tried to be increased.

Thus, ensemble learning methods, which are frequently used in the literature, are combined with fuzzy logic to increase the prediction success, and new studies have been added to the studies in this field. The designed model was first tested for the sales and cost data of online travel agencies. Since it was successful, it was applied to different problems in different sectors for its generalization feature. This approach was tested for the classification problem instead of the regression problem in later studies and proved successful.

1.1 Purpose of Thesis

The main objective of this thesis is to increase the prediction success rate by combining the ensemble learning methods, which are frequently used in the literature, and the fuzzy logic approach in a different way. This approach targets specifically datasets consisting of a cluster with different behavioural characteristics. Since items such as these can contain the characteristics of more than one cluster simultaneously, an approach is presented that makes the modelling and prediction process based on the cluster/clusters to which the prediction to be made in the single observation set is based. In the application, firstly, digital advertising and sales data

of online travel agencies were used. In later studies, user behaviours and digital advertising data of an application-based crossword puzzle game were used. In the first stage, the proposed approach was tested in regression problems, and in later studies, it was adapted for classification problems.

After a brief introduction chapter, I presented my thesis with four additional chapters. The following three chapters involve the presentation of my accepted papers in international SCI-E indexed journals and international books.

In Chapter 2, the traditional regression method and the proposed fuzzy clustering approach were used together, and the results were compared with the results of the traditional methods. Machine learning algorithms were applied directly to the dataset, which had been applied data preprocessing and feature engineering within the framework of traditional methods. Then the modelling study was carried out again after the hard clustering and soft clustering approaches were applied to the dataset.

In Chapter 3, the approach proposed in the first article is tested for solving a different problem with different sector data. In this study, it has been tried to predict the lifetime value of the customers by using the game data and session information of the users of a mobile crossword puzzle game published in more than thirty languages and more than thirty countries. Ensemble learning algorithms, which were not used in the first article, were examined in more depth and focused on algorithms that could achieve higher prediction success rates when used together with fuzzy clustering. Different hyperparameter combinations of Catboost, Extreme Gradient Boosting and Light Gradient Boosting algorithms, which are seen in the literature to be generally more successful than traditional machine learning algorithms, were tested separately for each cluster after the clustering phase performed with the Fuzzy C-Means clustering algorithm. The prediction values of the three most successful of these combinations were weighted to be inversely proportional to the error rates, and the error rates of the resulting predictions were compared with the results of other model-parameter combinations.

In Chapter 4, customer retention rate prediction was carried out with a different dataset collected in the gaming industry. Unlike the first two studies, in this study, a classification problem was tried to be solved with the proposed method; at the same

time, different cluster initial parameters and different fuzziness parameters were tested. The aim is to obtain a more optimal clustering in the Fuzzy C-Means clustering approach, and the clustering process was the most successful combination. Since the nature of the problem is a classification problem, the prediction was carried out by weighting the accuracy results instead of the error rates of the algorithms at the stage of combining the results of the algorithm-parameter combinations.

In Chapter 5, the conclusions are obtained from the proposed method and suggest some future research works.

2. CLICK AND SALES PREDICTION FOR OTAS' DIGITAL ADVERTISEMENTS: FUZZY CLUSTERING BASED APPROACH[1]

2.1 Introduction

The economic advantage of Web ads depends on whether the users click on the ad. Clicking the advertisement lets Internet companies recognize the most appropriate ads for each client and improve the customer experience. More precisely, one of the most important measurements utilized to determine the trading value of an ad is the click-through rate (CTR), which is the calculated sum of clicks divided by the sum of impression[5]. For search advertising, the CTR is used to rate advertisements and to calculate clicks[22]. The impression is a term which refers to the point where a visitor sees the ad once. Getting more CTR influences pay-per-click (PPC) performance as it legitimately prompts how much advertisers will pay for every click [19]. Pay-per-click advertisement is an auction-based system that usually appears to be the highest bidder in the most visible place in the ads.

Once the advertisement is clicked on, the advertising platform charges the bid amount for each click to the advertiser. Additionally, predicting each hotel's daily sales have demonstrated to be a difficult errand in light of the elements and unpredictability of the booking procedure. Bookings are influenced by numerous elements, for example, regularity, bunch bookings, occasions, lodging types, and events in the hotel, likewise, offering accomplishment in a serious situation. Joining these variables are significant for the success level of the prediction.

In this study, it is aimed to predict both the number of impressions and the click-through rate measurements of hotel ads and booking revenue, which is the aggregate of hotel reservation amounts online travel agency (OTA). To get a reservation from

[1] This chapter is based on the paper "Click and Sales Prediction for OTAs' Digital Advertisements: Fuzzy Clustering Based Approach". Tekin, A.T., Çebi, F. 2020. *Journal of Intelligent & Fuzzy Systems* 39(5), 6619-6627.

these engines, OTAs give digital advertisements to meta-search advertising channels with cost-per-click and cost-per-acquisition model. In our model, we focused on cost-per-click advertisement model. Precisely predicting the number of clicks for each hotel will, therefore, become compellingly important for online travel agencies to adjust their daily, weekly and monthly advertising spending plans and build their income models.

In this paper, we also present a relative overview of the efficiency measurements of the most advanced level prediction algorithms available in Opensource libraries, such as Random Forest, Gradient Boosting, eXtreme Gradient Boosting, and Support Vector Regression. Similarly, the performance information provided by a performance report produced by day-to-day metasearch advertising platforms is enhanced by some public information, such as currency, mete-orological outlook, public holiday data for each nation, and so on, which may be relevant to the advertising performance of the hotels.

Also, the clustering-based and regression-based approach are combined for improving CTR and Impression prediction results. In this method, clustering algorithms K-Means and Fuzzy C-Means are applied to data, then XGBoost, Gradient Boosting, and Random Forest algorithms are applied to each cluster and the overall results of the regression are evaluated. The results show that clustering the hotels and training them separately are improved the prediction results.

In the modelling section, some of the popular regression algorithms which are used in CTR, impression and sales prediction are described. Also, clustering algorithms which K-Means and Fuzzy C-Means are applied to data for clustering hotels according to their behaviour are described in the modelling section.

In the methodology section, applied techniques for data preprocessing and feature engineering steps before machine learning algorithms are applied are clearly explained. Also, features which are used in the modelling and regression algorithms' performance results are briefly explained in this section. Distinguishing the hotels which have different behaviour from the dataset, clustering algorithms are utilized then regression models are applied to data with combining approach and this process' results are shown in clustering-based and regression approach section.

2.2 Modelling

Models of regression are popular in machine learning and are used for predicting numerical target variables. There are many literature studies aimed at predicting digital advertisement impression count, click count and CTR level[18,19]. We used support vector regression (SVR), random forest, extreme gradient boosting (XGBoost), AdaBoost, gradient boosting, and deep neural network in this research to predict numbers of interactions, CTR rates, and volumes of sales that were effectively implemented for many regression tasks. In this section, those algorithms are briefly explained.

SVR is an open-source machine learning algorithm and the regression variant of the support vector machine. SVR agrees that modelling of non-linear regression problems is very successful[8]. In any case, anomalies or noise happen unavoidably for different reasons in the dataset, for example, numerical issues, changes in framework conduct, incorrect estimations, and inspired distorting. To decrease their negative impact, a few anomalies or noise are expelled legitimately, and others are recognized by outlier detection methodologies. If these anomalies or noise in learning progress are not adequately detected, they may decrease SVR's strength and performance. Along these lines, reducing the impact of anomalies or noise on the prediction stage is one of the primary objectives of SVR [33].

AdaBoost is also an open-source machine learning algorithm adapted as the first fruitful boosting algorithm. Though pro-positioned as an ensemble learning method for classification tasks, it was subsequently adapted to the regression tasks[20]. AdaBoost focuses on the observations misclassified by the former learner, expanding the observations of these perceptions; henceforth, changing the information distribution adaptively. As an exceptionally powerful ensemble learning technique, AdaBoost can get very high classification precision and is periodically less defenceless to the issue of over-fitting. Notwithstanding, AdaBoost includes the age of various base learners, representing a challenge to capacity resources [34].

Random forest is yet another machine learning algorithm depends on a combination of predictions from many decision trees. The idea behind those approaches to the ensemble is to build a strong solitary model on multiple frail models. There are many effective applications for various types of problems concerning machine

learning[12,21]. The model of random forest regression is an improved version of the CART methodology and may offer better results for prediction. The training phase of R.F. is to build different decision trees. Each tree in R.F. is devel-oped with a randomized subset of indicators and thus is called 'random' forest. It is an ensemble method that combines all of the generated decision trees using a calculation called bagging or bootstrap aggregation. Bagging is a technique proposed by Breiman[35] and can be used with numerous regression strategies to reduce the prediction-related change, thus enhancing the prediction performance. Randomly selected features build R.F. for each decision tree or randomly selected samples build R.F. for each decision tree. Randomly gathered observation data process is called bootstrapping [36].

XGBoost is an as of late proposed machine learning algorithm that is an adaptable machine learning approach dependent on a boosting form. It is getting progressively well known because of its prevalence over many machine learning algorithms in a few machine learning quests [6]. For example, in [16], XGBoost is increasingly successful in predicting a bicycle station's hourly requests over cutting-edge strategies. The most important factor for the prominence of XGBoost and its prosperity is its adaptability to all tasks of machine learning. The architecture runs much faster than current well-known regression algorithms on a solitary server and it responds to settings that are circulated or limited by memory[11]. XGBoost's scalability is attributed to a few major methodologies and advances in algorithms[13,7,16]. XGBoost is an improved GBDT (Gradient Boosting Decision Trees) algorithm that includes numerous decision trees and is used regularly in the classification and regression field. In any case, XGBoost varies in some respects from GBDT. To begin with, the GBDT algorithm only uses the Taylor development of the first order, while XGBoost includes a Taylor extension of the loss function for the second request. Second, normalization is used in the objective function to forestall over-fitting and decrease the complexity of the model [37].

To advance the hyperparameters of all machine learning algorithms utilized in this research, we used grid searching. Sixty per cent of the examples are used in this study for training, twenty per cent for validation and the remaining twenty per cent for testing. The data was shuffled during the splitting process, and use was made of the sci-kit library's data split module.

We also used an alternative type of non-parametric machine learning algorithms to allow the process of enriching data to be involved in predicting hotel sales. One such technique is the tree-based algorithms that combine numerous vulnerable learners to obtain a generalizable lone model. Extreme gradient boosting (XGBoost)[10] is a machine learning algorithm, which has since become famous among data scientists because of its notoriety in many machine learning com-petitions[15,17,23]. XGBoost[6] contains additional regularization parameters that control the size and shape of the trees, making predictions stronger and better suited to the algorithm. Ultimately, random forest, gradient boosting, and extreme gradient boosting (XGBoost) algorithms from tree-based algorithms were selected to be implemented in our research, as high accuracy appeared to be achieved on various regression problems[9,13,14].

The hotels in the dataset have different behaviours in terms of some performance metrics. Some of them have both impressions, click and sales information according to performance report data, while some of them don't even get impressions. Because of this difference, hotels are clustered according to their features in the observations in the dataset. For the clustering process, K-Means and Fuzzy C-Means algorithms are selected.

K-Means is a clustering algorithm that is a powerful and traditional method in clustering. The purpose of K-Means is minimizing the sum of squared error in a given dimensional space with selected k initial seeds. In terms of its working principles and being easy to interpret, K-Means is a clustering algorithm which is widely used clustering algorithm [24,25,26,27].

Fuzzy C-Means is also a clustering algorithm that proposes a membership function for each variable and it calculates the association level of each variable and each cluster. So, this method allows multiple memberships for each variable [28]. Fuzzy C-Means algorithm is defined by Dunn and developed by Bezdek [28,29]. This method is widely used in several studies and several wide varieties of substantive areas.

For performance evaluation in our CTR, impression and sales prediction, we used some performance metrics which are popularly used in regression prediction in the

literature. These metrics are Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and R-Squared (R2) value. Explanation of these metrics is given in below.

MAE: The MAE is defined by Equation 2.1. MAE averages the absolute differences between predicted and actual values. The smaller amount of MAE means that prediction is more accurate.

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{2.1}$$

In this equation, n denotes the number of samples, then it is calculated by the sum of absolute values of differences between actual and predicted values.

RMSE: This metric is very similar to MAE. It is also used to measure the difference between predicted and actual values. The difference between MAE and RMSE is because of RMSE gets square of error, it punishes more the larger differences in predicted and actual values. RMSE is defined by Equation 2.2.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{2.2}$$

R^2: This metric is another measurement of how intently predicted and actual values match each other. R^2 is a score between 0 and 1. When this score is getting closer to 1, it means that prediction accuracy is getting better. It is calculated by subtracting division of explained variation to total variation from 1. R^2 is defined by Equation 2.3.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \tag{2.3}$$

2.3 Proposed Methodology

2.3.1 Regression based approach

In our click prediction stage; Hotel impression (shortly will be alluded to as impression) is the number of impressions got for a hotel. When the client sees the hotel, it is listed on the search result page of a meta-search engine for a hotel. It is a major predictor of a hotel's reputation and can be used to measure a specific hotel's traffic capability. This number is synonymous with the hotel's marketing potential and reputation. The meanings of the significant measurements utilized in this paper given underneath:

Click: The number of clicks the meta-search advertising platform checks.

Click-Through-Rate (CTR): This ratio shows the sum of total clicks divided by total impressions sum.

CPC: The OTA will pay the amount of money for each click to the meta-search engine for the respective date.

Cost: It is the product of the total number of clicks received by the hotel for the date and cost-per-click value corresponding.

Initially, data cleaning techniques were administered to data. The features that are hotel URL, hotel name, etc. that could not be used for machine learning algorithms have been wiped out of the data for this reason. Also, duplicate rows that ought not to be on the data were wiped out, like more than one row, which is about a hotel from that day. Then, steps to enrich the data were applied to the dataset. In the data enrichment step; certain features have been added to the dataset with shifted and rolling average. Similarly, as indicated by their location, the hotels were sorted as city or summer hotels.

We created a new feature named “hotel_type” to speak with this category of the hotel to data such as a resort or city hotel. Given the importance of coming public holidays in predicting the potential number of clicks and reservations, the holiday period and how many days up to the start of the holiday are implemented as new features in the dataset. The price of the hotel in the corresponding date and position information (position number of the OTA advertisement) is also added as new features for each hotel in the meta-search bidding engine, which additionally incorporates the prices and positioning information of the closest competitors of the OTA, which are so vital to the competitiveness of the OTA sources.

Missing values in the dataset were filled in after the enrichment process of the data. The OTA performance report included a few missing values which can be filled in by using some of the statistical methods suggested in the literature. For example, if the value of the "click" variable is 0, and the cost information for this hotel is missing, the cost is filled with 0 because it is accepted that in such a situation the hotel in question did not take any clicks on the corresponding date if the click is 0, the marketing cost for that hotel will be 0. Missing values in hotel-related characteristics, such as stars, the hotel's rating is filled with the column average. The

categorical values that speak to a hotel property (e.g., the city information of the hotel) are filled with the most frequent data point in that hotel area. Ordinal categorical variables, which are string values like the booking value index, are converted to numeric values 1 through 5.

In our issue, the OTA will provide the metasearch engine performance reports which have the most important features for our model on the next calendar day. The average values of some of the features are shifted and rolled average as 3,7 and 30-day values applied to the training set to address this problem and also to use the relevant sequential features in the predictive function. Additionally, the data is added to the training set on the day of the week, as it tends to be a significant click amount pointer for explicit hotels. Furthermore, the value of the bid, click and profit for each hotel are added to the dataset; both the last values from the previous day and the values from the same weekday from a week earlier. The hotel's prices over the last ten days are also added as separate columns to catch the changing pricing patterns. The data set that was used in prediction consisting of nearly 220 features and more than 800,000 samples was obtained as a result of the data preprocessing steps mentioned above. The modelling features are given in Table 2.1.

Table 2.1: Base features used in CTR and impression prediction.

Feature	Description	Type	Range	Is Categorical	Missing Value Rate
impression	Number of impression recieved for hotel in corresponding date.	Numerical	[0,24300]	0	0%
clicks	Number of clicks recieved for hotel in corresponding date.	Numerical	[0,1208]	0	0%
beat	The ratio of in how many impression, hotel beats to competitor price	Numerical	[0,1]	0	2.4%
meet	The ratio of in how many impression, hotel meets with competitor price	Numerical	[0,1]	0	2.4%
lose	The ratio of in how many impression, hotel loses according to competitor price	Numerical	[0,1]	0	2.4%

Table 2.1 (continued): Base features used in CTR and impression prediction.

unavailability	The ratio of in how many impression, hotel is unavailable in related search criteria	Numerical	[0,1]	0	2.4%
top_pos_ share	The ratio of in how many impression, hotel ranks in the top position	Numerical	[0,1]	0	2.1%
stars	It indicates how many stars hotel has	Numerical	[0,5]	0	1%
rating	Rating value of the hotel in metasearch platform	Numerical	[0,100]	0	17%
ctr	Click through rate	Numerical	[0,1]	0	0%
max potential	Maximum impression hotel can get	Numerical	[0,1]	0	0%
impr_share	Percantage of impression company's offer listed in the search result	Numerical	[0,1]	0	0%
hotel type	It indicates hotel is resort hotel or not	Categorical	Summer, City	1	2.4%
bid	Cost per click amount	Numerical	[0, 0.75]	0	0%
city	In which city hotel is located	Categorical	-	1	1%
booking value index	Estimated booking average category according to it's conversion level	Categorical	5 different values	1	3.5%
log_date	The date which shows the corresonding date	Date	-	0	0%

Table 2.2: Algortihm's results for CTR prediction.

CTR	**Algorithms' Result**				
Algorithms	**R^2**	**RMSE**	**MAE**	**CV Mean R^2**	**Sum Success**
Random Forest	0.55	0.046	0.022	0.52	0.97
Gradient Boosting	0.57	0.045	0.021	0.58	0.99
AdaBoost	0.3	0.197	0.17	0.12	0.35
SVR (kernel='rbf')	0.25	0.098	0.083	-	0.47
XGBoost	0.61	0.045	0.02	0.59	0.98

Table 2.3: Algortihm's results for impression prediction.

Impression Algorithms	Algorithms' Result			
	R^2	RMSE	MAE	SumSuccess
Random Forest	0.5	37.37	16.87	0.93
Gradient Boosting	0.63	32.16	15.94	0.97
AdaBoost	0.4	490.79	383.21	0.08
SVR (kernel='rbf')	0.35	105.12	67.84	0.44
XGBoost	0.81	27.84	13.54	0.95

Table 2.4: Algortihm's results for click prediction by predicted impression multiplied by predicted CTR.

Algorithms	Algorithms' Result				
	R^2	RMSE	MAE	CV Mean R^2	Sum Success
Random Forest	0.80	595.25	260.92	0.81	0.98
Gradient Boosting	0.80	596.35	268.79	0.79	0.98
AdaBoost	0.35	1457.39	1236.26	0.20	0.50
SVR (kernel='rbf')	0.27	1423.74	657.33	-	-
XGBoost	0.84	637.40	274.17	0.84	0.99

Tables 2.2, 2.3 and 2.4 show prediction results obtained by taking care of all the features as input to the machine learning algorithms. The results show that for both CTR and impression prediction, XGBoost performs broadly superior to any other machine learning algorithm. The most noteworthy R-Squared value acquired when predicting CTR and impression values based on individual hotels is 0.61 and 0.84, separately, both accomplished by XGBoost. After XGBoost the other two tree-based algorithms, Random Forest and Gradient Boosting are then placed. The results show that SVR and AdaBoost don't produce generalizable models on this task. In the click the prediction stage; the most noteworthy R-Squared value of 0.81 is also gotten with the XGBoost algorithm. The results showed that impression prediction is more successful than CTR prediction.

Sum Success ratio which indicates the ratio for the sum of actual clicks divided by the sum of predicted clicks were also added as a performance metric for each algorithm. This ratio is also another important metrics for algorithms' performance. Because it also represents nearly the sum of the marketing cost for the next day. The tree-based ensemble methods are seen to offer comparable results for this problem, which overall exceeds ninety-five per cent.

Revenue forecasting with the marketing cost is crucial for top management for companies. So, sales prediction is also important for OTAs. In our sales prediction, predicted click amounts which were from the previous study. Because the sales

amount is positively correlated with marketing cost. So, for the sales prediction, with the sliding window method, we tried to predict the next day's sales number with predicted click values. For this purpose, additional features provided by OTA performance report are generated as moving averages and shifted values from specific features. These features include moving averages and 3, 7, 15, 30 and 45-day standard deviations of the original features. Also, the day-to-day sliding windows approach has been updated for only the features that are correlated to the target variable. The most correlated features were the slided values of the target variable, which were features of the sliding windows of the previous days. Analyzing the correlation showed that the total number of nights and rooms was directly related to sales. Such features as sliding features of the information from past days have been applied to the dataset.

Likewise, additional features with daily moving feature averages of their ten days were added to the dataset. Some of these features are the cost of opportunity per click, clicks, average booking value, bid, cost, gross revenue, the number of people who stayed at the hotel, respectively. In addition, the moving sums of those values from the previous 3, 7, 15, 30, 45 days are included in the dataset. Also, we were able to get the number of bookings that have the breakdown of completed, cancelled, and pre-booking details from the booking data. These numbers have been added to data in previous periods with their moving averages and standard deviations. But, because of the reservation data depends on the target variable, actual values of reservation data is not added to dataset directly. These data were added to dataset as previous days' moving averages and standard deviations of previous days.

After all of these steps, one-hot encoding method was applied to dataset for categorical variables. Then, some of the features which are related to the next day's target variable were dropped from the dataset because of avoiding the bias-variance trade-off.

The dataset which was obtained after all of these data preprocessing and feature enrichment steps contains nearly 375000 rows and 315 features which belong to the dates which are used in CTR and impression prediction dataset. The additional features which are used in the sales prediction but not in the CTR and click prediction are given in Table 2.5.

Table 2.5: The additional features which are used in the sales prediction.

Feature	Description	Type	Range	Is Categorical	Missing Value Rate
Total Night	How many nights customer will stay in that reservation	Numerical	[1,28]	0	0%
Total Rooms	How many rooms customer booked in that reservation	Numerical	[1,12]	0	0%
Reservation Revenue	How much customer paid for that reservation	Numerical	[10,19800]	0	0%
Base Price	The ratio of in how many impression, hotel meets with competitor price	Numerical	[9,17800]	0	0%
Commission Percentage	Reservation Revenue – Base Price	Numerical	[0.10,0.20]	0	0%
Reservation Status	It indicates that the reservation is pre-reservation, cancelled or completed reservation	Categorical	-	1	0%

In the final step, eXtreme Gradient Boosting, Gradient Boosting Machines, Random Forest, Generalized Linear Models and Deep Neural Network was applied to the dataset which is mentioned above. Sum of success criteria of these machine learning algorithms is given Table 2.6.

Table 2.6: Algorithms' results for sales prediction.

Model	R^2	MSE	MAE
eXtreme Gradient Boosting	0.57	2267.02	9.45
Gradient Boosting Machines	0.56	2159.98	9.54
Random Forest	0.57	2121.13	9.40
Generalized Linear Models	0.44	2760.8	16.18
Deep Neural Network	0.51	2442.98	15.15

2.3.2 Clustering based and regression based approach

The hotels in the dataset have different behaviours from each other and also, some of the hotels which are in the dataset are affected by seasonality.

So, instead of keeping these hotels in the same dataset, hotels are clustered and it was aimed to improve performance metrics. Due to this reason, K-Means and Fuzzy C-Means clustering algorithms were applied to the dataset. The dataset consists of nearly ten thousand unique hotels' data which are formed from CTR, impression and sales prediction's dataset. For setting k initial seed in the clustering, Silhouette scores for k values which are 3, 5, 7 and 9 were evaluated. In each method, k values are set to five, which has greater Silhouette score than other k initial seeds and for

successful clustering validity index values were analyzed. The number of sets for clusters is given in Table 2.7.

Table 2.7: Number of set for clusters.

Clustering Algorithm	C1	C2	C3	C4	C5
K-Means	2350	1640	4229	532	1141
Fuzzy C-Means	2634	1921	4597	943	1367

In K-Means clustering, Silhouette coefficient is a widely used metric for validation criteria. Silhouette width presents the difference between the intra clusters tightness and separation from the inter-clusters [30,31]. In Fuzzy C-Means clustering, Fuzzy Partition Coefficient (FPC) is a popular index which is used for measuring the quality level of clustering [32]. FPC score changes between 0 and 1 and the maximum value indicating that the best clustering quality.

Silhouette score of K-Means clustering was nearly 0.9 and it indicated that clustering of the hotels was successful. On the other hand, the FPC score for Fuzzy C-Means was 0.96 and also it indicated that this clustering process was successful as well. Due to the working principles of Fuzzy C-Means clustering algorithm, the 0.25 threshold point was set after calculating the relationship level of each hotel with the relevant cluster. So, a hotel can be a member of more than one cluster according to its relationship level.

In the second stage of this approach, each algorithm and hyper-parameters were tested for each cluster in both clustering algorithms' result set with XGBoost, Random Forest and Gradient Boosting algorithm which are the most three successful algorithms in impression and CTR prediction. The results of impression and CTR prediction in each cluster and overall performance of full dataset are given in Table 2.8, Table 2.9 and Table 2.10.

Table 2.8: Comparison of clustering algorithms for CTR prediction.

Clustering Algorithm	Regression Algorithm	RSQ	RMSE	MAE	Sum Success
K-Means	XGBoost	0.64	0.039	0.018	0.98
K-Means	GradientBoosting	0.58	0.041	0.019	0.98
K-Means	Random Forest	0.57	0.042	0.021	0.97
Fuzzy C-Means	XGBoost	0.67	0.034	0.016	0.99
Fuzzy C-Means	GradientBoosting	0.60	0.038	0.017	0.98
Fuzzy C-Means	Random Forest	0.60	0.039	0.02	0.98

Table 2.9: Clustering algorithms and regression algorithms' results for impression prediction.

Clustering Algorithm	Regression Algorithm	RSQ	RMSE	MAE	Sum Success
K-Means	XGBoost	0.85	610.23	271.2	0.99
K-Means	GradientBoosting	0.82	588.51	264.8	0.97
K-Means	Random Forest	0.81	592.44	260.1	0.94
Fuzzy C-Means	XGBoost	0.86	601.86	264.9	0.99
Fuzzy C-Means	GradientBoosting	0.82	589.32	264.9	0.97
Fuzzy C-Means	Random Forest	0.81	594.23	261.5	0.96

Table 2.10: Clustering algorithms and regression algorithms' results for click prediction by predicted impression multiplied by predicted CTR.

Clustering Algorithm	Regression Algorithm	RSQ	RMSE	MAE	Sum Success
K-Means	XGBoost	0.83	26.01	13.44	0.97
K-Means	GradientBoosting	0.66	30.88	15.64	0.96
K-Means	Random Forest	0.51	36.89	16.01	0.95
Fuzzy C-Means	XGBoost	0.86	24.12	12.62	0.98
Fuzzy C-Means	GradientBoosting	0.66	30.92	15.77	0.98
Fuzzy C-Means	Random Forest	0.54	34.51	13.93	0.95

2.4 Conclusion

In this study, we intended to predict the number of clicks every hotel will get in the meta-search advertisement platform for the following day using historical data for the click prediction. To this end, we first used numerous data preprocessing techniques and arranged the dataset containing the moving average and standard deviations of the original features, then applied some of the feature selection methods proposed in the literature to reduce the feature dimension and used the final data set selected in the feature selection process and we gave a set of machine learning algorithms as input to this dataset. The primary commitment of this paper is acquiring the predicted clicks depends on the prediction of the CTR and the values of hotel impression. Predicted click amount of each hotel for the next day was obtained by multiplying predicted CTR value and predicted impression value for the corresponding hotel.

The results show that 0.81 was the highest R-Squared value acquired by multiplying CTR and impression, which eXtreme Gradient Boosting obtains. The other criterion of success, which can be considered a total success, depends on the contrast between the actual amount of click divided by the predicted amount of click. This value is an

indicator of the total marketing costs for a particular day concerned. We achieved a 95 per cent Sum-Success criterion, showing the viability of the features wiped out from the initial dataset.

For improving these results, hotels were clustered via K-Means and Fuzzy C-Means clustering algorithms and results were determined. In Fuzzy C-Means clustering algorithm, 0.96 FPC score was obtained which means hotels were clustered successfully. Fuzzy C-Means clustering algorithm improved R2 value of multiplying CTR and impression from 0.81 to 0.86 and sum-success validation criteria from 0.95 to 0.98.

Fuzzy approach prevented that decreasing hotel numbers' in the clusters so, more training data for clusters are used in training than K-Means clustering. This approach shows us if the data is limited after clustering operation, Fuzzy C-Means can be used because of its membership functionality.

For the sales prediction, we also generated a dataset which contains performance metrics based on the metasearch engine performance report and moving averages and standard deviations of some of the performance metrics. Also, predicted click values which were obtained from CTR and impression prediction. These predictions were from our previous prediction stage. Different predictive algorithms including gradient boosting, XGboost, random forest, generalized linear model, and deep neural network were applied after the data preprocessing and feature enrichment stage to predict the amount of sales for the corresponding hotel next day. The results showed that the predicted amount of clicks was crucial in predicting the amount of sales. As a nature of this problem, click amount is strongly correlated with marketing cost and marketing cost is strongly correlated with the sales amount. The results also showed that adding target variable related features to the dataset as moving averages and standard deviations of these features improved the model success. For evaluating the success level of machine learning algorithms, we can say that tree-based algorithms performed better than the other algorithms. We should also say that XGboost is slightly better than Gradient Boosting, but according to generalized linear models and deep neural network, it is ahead.

3. CUSTOMER LIFETIME VALUE PREDICTION FOR GAMING INDUSTRY: FUZZY CLUSTERING BASED APPROACH[2]

3.1 Introduction

Although the use of machine learning algorithms has increased, especially in the last period, their entry into the literature dates back to quite old times. With the development of technology, the number of operations that computers can perform per second increases, and therefore the frequency of using machine learning approaches in problem-solving also increases. With machine learning approaches, both academic and private-sector problems are tried to be solved and their impact on the solution of these problems is increasing day by day. The solution to modern world problems will become easier when data, whose size is increasing day by day in the contemporary world, can be processed or interpreted.

Machine learning problems that are tried to be solved, especially in the private sector, are generally on production, sales, expenditure and customer movements, etc. As a result, whatever the predictions are, the aim is to predict the unpredictable events that will occur in the future and contribute to the decision-making process. For this purpose, companies plan their activities according to the scenarios that may arise in the future and have information about what awaits them. In prediction problems, regardless of the type of estimation, the purpose is to perform the prediction with the lowest error rate.

Many techniques have been carried out in the literature to meet this purpose, and one of the most widely used strategies is Ensemble Learning. This method aims to combine multiple algorithms to close their vulnerabilities and decrease the prediction error rate. The success of this method has been proven in the literature, regardless of whether the problem is classification or regression.

---

[2] This chapter is based on the paper "Customer Lifetime Value Prediction for Gaming Industry: Fuzzy Clustering Based Approach". Tekin, A.T., Kaya, T. ,Çebi, F. 2021. *Journal of Intelligent & Fuzzy Systems 42*(1), 87-96.

The data set can be used directly to solve classification and regression type problems, which are sub-branches of machine learning or can be used as separate parts after being divided into parts with the clustering approach. Again, the aim is to apply machine learning algorithms to input datasets that are similar to each other to reduce the error rate.

The theory of fuzzy logic is also a technique often used for problems of machine learning. One of the primary purposes of using fuzzy logic in machine learning problems is that the problems trying to be solved do not consist of exact answers such as zero and one. We are attempting to illustrate the solution to various issues in our present life with values between zero and one.

For the bagging and stacking of ensemble approaches, fuzzy logic was used in this research. The purpose of this application is to cluster the set of observations of related behavioural features. The products or users that are tried to be predicted by machine learning have different characteristics. Modeling these products or users together can negatively affect the success of the model. When trying to cluster a product or a user, the problem of associated with more than one cluster at the same time is encountered. Fuzzy clustering is an ideal approach to clear up this problem. In this way, the product or users can be modelled separately in each cluster they belong to, and thus the success rate of the model can be increased.

The paper addresses the literature review of ensemble learning and fuzzy methods of machine learning in Section 3.2. Our proposed approach and modeling data are illustrated in Section 3.3. Finally, the study's results are briefly outlined and the last segment presents potential work.

3.2 Literature Review

Ensemble learning is the technique used in the same problem to train more than one algorithm and to continue to solve the same problem. In comparison to the methods of individual machine learning algorithms[38], learning takes place in an attempt to combine to construct and use a set of hypotheses. Ensemble Learning's fundamental goal is to produce greater predictive efficiency than specific machine learning algorithms.

Usually, ensemble learning algorithms are made up of various algorithms called simple algorithms. Algorithms for ensemble learning have a higher generalization potential than simple algorithms. Ensemble Learning was first extended to problems with classification and then adapted to regression problems [20].

Unfortunately, the approaches applied to classification problems are often not valid for regression problems. [41]. In other words, a recent analysis used for classification in the Ensemble Learning Methods is not enough to provide a summary of the latest approaches to the regression problem. More than one model is generated as a working assumption and the samples to be measured are supplied to these models as inputs. Outputs are transferred into the voting system and the final step of the calculation is carried out. While in each cycle, the ensemble learning steps differ, they can typically be represented as follows[47].

The training data set consists of m = D = {(x1, y1), (x2, y2), ..., (xm, ym)}, k class tags are yi εY = {1, ..., k}, classification assume that the algorithm is denoted by L, the population size is set to n.

Step 1: D dataset can be used directly for training (Voting) or to build new Di datasets from D dataset (Bagging, Acceleration).

Step 2: The following process shall be repeated n times. During this replication, the same data set is trained with different Ci = Li (D) learning algorithms or different data sets are trained with the same Ci = L (Di) algorithm.

Step 3: The ordinary judgments of the classifiers shall be checked with test data collection.

Step 4: The output of each classifier is generated for a new sample x. yi = Ci (x).

Step 5: The results of n classifiers {C1, C2, ..., Cn} are merged.

$$C^*(x) = \arg max_{y \in Y} \sum_{i1C_i(x)=y}^{n} 1 \quad (3.1)$$

Many literature studies have shown that, compared to a single simple algorithm, ensemble learning algorithms have higher success rates in prediction[39,40]. Medes-Moreira et al.[41] have looked at current learning methods for regression since ensemble learning approaches to regression algorithms are different from classification problems.

In the literature, there are three effective approaches proposed to the ensemble learning phase of regression problems. These methods are: the Stochastic Gradient Boosting [42, 45], the Standard Bagging [35] and the hybrid version of these two strategies are the Bagging and Stochastic Gradient Boosting strategies. Bagging and Stochastic Gradient Boosting is often referred to as another form of MultiBoosting Approach [43]. If we look at these approaches in detail;

- Stochastic Gradient Boosting: With the bagging procedure, Breiman [35] argued that adding random sampling to the estimation procedures will positively affect the estimation process's performance. In the same period, random sampling was used in the Adaboost [44] method, which is another popular ensemble learning method. However, in the Adaboost method, if the learner algorithm that is based on does not support observation weights, this based learner algorithm is accepted as an approach that promotes observation weights, not a basic component [42]. According to the literature, the bagging method is based only on the variance in the dataset, while the Adaboost method is based on both variance and deviation. Stochastic Gradient Boosting incorporates both the boosting and bagging approach from ensemble learning approaches. Many small classification or regression trees are created sequentially based on the previous tree's loss function gradient. At each iteration step, a tree is created that provides an incremental improvement in the model built using a random sub-sample of the data set [45]. Using a subset rather than the entire data set improves both computation speed and prediction accuracy. This method is also sensitive to outliers in the data.
- Standard Bagging: The bagging method creates multiple versions of an estimator and then creates a combined estimator from these estimators [35]. The combined estimator is based on the voting system when estimating a class while estimating a numerical result, averaging the estimates created in different versions. In this method, while creating regression trees, a training set of the same size as the original data set is created based on the "bootstrap" approach. Some items can be left out in this training set, while some items can be used repeatedly. Breiman [35] stated that for the created bags to be effective, the observation data in the bag is unstable and depends on the rate of responding to changes in the training data.

- MultiBoosting: In this approach, it was emphasized that Adaboost and bagging approaches have separate effects and according to the training made on the original data set, both approaches have positive effects separately. Considering these effects, it is based on combining two different outcomes [43]. According to Webb [43], bagging is mainly aimed at reducing variance, while Adaboost is an approach to reduce both variance and deviation. However, according to Bauer and Kohavi [46], the bagging approach gives more positive results than Adaboost in reducing variance. Adaboost [44, 46] is a kind of ensemble algorithm like Stochastic Gradient Boosting, based on random sampling.

Other ensemble learning approaches recommended in the literature and used in machine learning problems, aside from these techniques, include voting and stacking. In the stacking process, several algorithms generated using different L1, L2 ... learning algorithms are merged. The S data set which does not consist of a single data set, is mapped with the vectors property (xi) and the groups (yi) of the vectors si = (xi, yi). In the first step, a group of simple classifiers C1, C2, ..., Cn is formed. Another meta-level algorithm, which combines the results of these classifiers, is trained in the next step. Validation or cross-validation approaches that remove an object are used to build this meta-level algorithm [48].

The decisions taken can be more easily interpreted in the voting method. To ensure the correctness and efficiency of decisions taken in the present life, more than one person is asked the same question. The majority of votes support the most answered answer, and its precision is more accurate. In the basic voting system, all the classifiers' votes are of equal weight [49]. The decisions taken by each classifier are combined and the mark with the most votes is chosen.

Decisions taken do not have to be applied with equivalent weights. If required, the results of learning algorithms on decision-making can be modified by adjusting the decision weights.

Although the voting system is used for classification problems, the same approach for regression problems is the average or weighted average. In this method; Calculations of learner regression algorithms may be averaged or their weights may differ by the target output parameter. This technique is called the weighted-average method of Ensemble Learning.

Fuzzy Logic [4, 51, 52] was discovered by Lofti Zadeh, professor of computer science, in 1965. In the 1950s, Professor Zadeh claimed that all real-world problems could be solved effectively by analytical or computer-based methods [53]. In 1964, he developed the "Fuzzy Set Theory", which has an essential position in literature. While this hypothesis has been questioned by some scholarly communities for its complexity, it is being used in many fields today.

The fuzzy cluster is described by a function that maps objects to the membership value of the cluster in the respective domain [53]. Fuzzy logic is a method for determining intermediate values between two standard measures, such as true/false or yes/no.

Machine learning algorithms are primarily intended to extract information from data and are used in traditional clustering methods, classification and correlation for this purpose [50]. As fuzzy set theory appears to produce more scalable results, machine learning approaches are common. Since fuzzy set theory can model missing and incorrect data as a function, it is used in various stages of machine learning, including data processing, feature engineering, and simulation.

Fuzzy logic appears as fuzzy classifiers in the literature and is also a widely used approach for classification problems. With numerical expressions, groups can also be specified. A fuzzy classification scheme can be described in this case by means of a simple law. There is a mixture of linguistic vector values to the left of each law that describes a particular class [54]. On the right side is the integer variable which represents the same class. An example rule set that describes this condition is given below [55].

If x1 is medium and x2 is small, THEN class: 1.

If x1 is medium and x2 is large, THEN class:2.

If x1 is large and x2 is small, THEN class: 2.

If x1 is small and x2 is large, THEN class: 3.

In the literature of cluster problems related to unattended learning problems, fuzzy logic theorems are also frequently found. Clustering algorithms, hard and flexible clustering[56], may typically be treated in two classes. Each finding in the test data set belongs to a single cluster within the hard clustering process. Besides, one item

will belong to more than one cluster in the flexible clustering process[50]. Flexible clustering is also used as fuzzy clustering in the literature. In the fuzzy clustering process, for each observation value, the membership level is calculated. This membership value is between zero and one.

There are several approaches to fuzzy clustering proposed in the literature. Fuzzy C Means Clustering [28, 29], Possibilistic C Means Clustering [57], Fuzzy Possibilistic C Means Clustering [58], and Possibilistic Fuzzy C Means Clustering [31,59] are some of the methods that have been proposed. The Fuzzy C Means Clustering, which is also used in this article, is the most popular. Dunn [29] first proposed this approach in 1973, and Bezdek [28] assembled it in 1981. This method is made up of two key measures.

- Measurement of cluster centres.
- Measure the distance of each observation set to these centres by Euclidian distance estimation process and assign it to the centres.

This algorithm assigns a membership value of zero to one to each of our outcomes. The degree of turbidity within the cluster is also calculated using the turbidity metric. As a result, if the observation value indicates that the individual is a member of more than one cluster at the same time, this condition is identified and its degree measured using this process.

We used an alternative form of a non-parametric machine learning algorithm in the machine learning stage to allow the data enrichment approach to forecasting the customer lifetime value. Tree-based algorithms, for example, aggregate a large number of weak learners to produce a generalizable lone model. Extreme gradient boosting (XGBoost) [10] is a machine learning algorithm that has grown in popularity among data scientists as a result of its success in many machine learning competitions [15, 17, 23]. Additional regularization parameters govern the size and shape of the trees in XGBoost [10], making predictions stronger and better suited to the algorithm. At the end of the day, the extreme gradient boosting (XGBoost) algorithm from tree-based algorithms was chosen to be used in our study because it appeared to achieve high precision on various regression issues [9, 13, 14]. XGBoost applies a regularization principle to increase the tree's classification function's size and make it more repeatable. Regularization also helps predict feature value, which

is critical in big data problems [10]. Equation 3.2 describes the estimated output of XGBoost.

$$Z=F(x_i)=\sum Tt=1ft(x_i) \quad (3.2)$$

where xi denotes the explanatory variables, and ft(xi) is the output function of each tree.

Catboost is also a boosting type ensemble algorithm which is very popular in the literature lately. Catboost was proposed by Prokhorenkova et al. [60] in 2018. This algorithm is based on the ordering principle to solve machine learning problems. It is a kind of modification of the standard gradient boosting algorithm which avoids target leakage. It is created for processing categorical features, which are commonly used in machine learning problems. It is an implementation of gradient boosting, which uses binary decision trees as base predictors. According to Dorogush [65], Catboost uses random permutations to adjust leaf values while selecting the tree structure to avoid overfitting problems caused by gradient boosting algorithms. Equation 3.3 shows the estimated output of Catboost.

$$Z=H(xi)=\sum Jj=1cj1\{x\in Rj\} \quad (3.3)$$

H(xi) denotes a decision tree function of the explanatory variables xi, and Rj is the disjoint region corresponding to the leaves of the tree [60].

LightGBM is one of the most successful boosting type ensemble algorithms. In LightGBM [61], at each point of gradient boosting, categorical features are converted to gradient statistics. This method will significantly increase calculation time. It measures statistics for each categorical value at each stage and memory usage to store the category belongs to which node depends on a categorical function at each break. However, it provides valuable details for constructing a tree. LightGBM combines tail categories to one cluster to solve this problem, thereby missing part of the results. Also, the authors contend that converting categorical features with high cardinality to numerical features is even better [62]. LightGBM expands the decision tree vertically, while others extend it horizontally compared to base gradient boosting techniques or Extreme Gradient Boosting. This feature allows LightGBM to process large amounts of data efficiently [66].

In the performance appraisal stage of machine learning algorithms, we choose to use the Root Mean Squared Error (RMSE) metric, which is a standard metric in the

literature on machine learning. To measure the difference between real and expected values, this metric is used. RMSE fetches the error square. So, the greater gaps between expected and real values are more punishable. RMSE is defined by Equation 3.4.

$$RMSE=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}} \quad (3.4)$$

3.3 Proposed Methodology and Modelling

The dataset used in this study is related to a crossword puzzle game published in Google Play Store and App Store. This dataset consists of users' first 24 hours of gameplay data and session information. Also, we used users' campaign information related to their attribution to the game. This study aims to predict each users' customer lifetime value for three months from the attribution time. This prediction indicates the revenue which will be acquired from customers after they interact with the in-app advertisements. This prediction is crucial for companies because it shows that users will bring revenue more than acquisition cost or not and companies' future strategies for getting more income from that customer. Base features which are used in this study are shown in Table 3.1. The dataset which is used in this study consists of 22 base features and 598478 rows. Each row in the dataset is related to a user's first-day gameplay and session information in an aggregated format. Also, all of these users have completed the three months after the attribution date.

Table 3.1: Base features used in lifetime value prediction.

Feature	Description	Type	Range	Is Categorical	Missing Value Rate
Session_ cnt	Number of sessions started for user in first 24 hour.	Numerical	[1,1391]	0	0%
Session_ length	Duration of sessions for user in first 24 hour	Numerical	[0,9016]	0	0%
App version	Application version of the game which users played in first 24 hour.	Categorical	33 different values	1	3.03%
language	The language information which users most played in first 24 hour.	Categorical	31different values	1	3.22%
Max_level _no	Maximum level number which users completed in first 24 hour	Numerical	[0,450]	0	0%
Gameplay _duration	Total Duration of levels which users completed in first 24 hour	Numerical	[0,8222]	0	0%

Table 3.1 (continued): Base features used in lifetime value prediction.

Bonus_cnt	The number of bonus which users used for completing levels in first 24 hour.	Numerical	[0,4977]	0	0%
Hint1_cnt	The number of first type hint which users used for completing levels in first 24 hour.	Numerical	[0,1222]	0	0%
Hint2_cnt	The number of second type hint which users used for completing levels in first 24 hour.	Numerical	[0,609]	0	0%
Hint3_cnt	The number of third type hint which users used for completing levels in first 24 hour.	Numerical	[0,1212]	0	0%
Repeat_ cnt	The number of repeating the levels which users used for completing levels in first 24 hour.	Numerical	[0,3559]	0	0%
Claim_ gold	It indicates user claims gold or not in first 24 hour.	Categorical	Yes or no	1	0%
Banner_ cnt	The number of banner type advertisements which users display in first 24 hour.	Numerical	[0,1465]	0	2.4%
Interstitial _cnt	The number of interstital type advertisements which users display in first 24 hour.	Numerical	[0, 411]	0	2.6%
Rewarded _video_ cnt	The number of rewarded video type advertisements which users display in first 24 hour.	Numerical	[0, 553]	0	3.2%
revenue	Revenue amount which is acquired from users in first 24 hour.	Numerical	[0,9.3]	0	7.8%
Shop Revenue	In App Purchase Revenue amount which is acquired from users in first 24 hour.	Numerical	[0,75]	0	98.2%
Played with friend	It indicates user plays with multiplayer option or not.	Categorical	Yes or no	1	0%
Multiplayer rating	The rating of users in leaderboard if user plays with multiplayer option	Numerical	[0,1053]	0	0%
Multi install	It indicates user deletes the game and re install the game.	Categorical	Yes or no	1	0%
Os_version	Operating system version of user which uses in his/her phone	Categorical	52 different values	1	0%

Initially, methods of data cleaning were administered to the dataset. For this reason, features that could not be used for machine learning algorithms were abolished from the dataset. These features were device id, session id, session start and end time, etc. After that, steps to enrich the data collection were applied to the dataset. In the data enrichment step, certain functions with session-related information are maximum session length, median session length, average session length, session per day information, which are session-related information and campaign information related to their attribution to the game added to the data set.

We obtained 30 features and 598478 rows as a dataset used in the modelling section after the feature engineering and data enrichment phase.

After the enrichment phase of the data, missing values in the dataset were filled in. The gaming info, app version and language, etc., contained a few missing values that can be filled in using any of the literature's suggested statistical methods. For example, if the shop's revenue is missing, it is filled with 0 for that user because it is accepted that there is no in-app purchase event for that user in the first 24 hours. Missing values in gameplay data, such as app version, language are filled with the most frequent value. The categorical values that speak to a user's gameplay data are string values converted to numerical values with one hot encoding technique. Also, the min-max scaling technique is applied to the dataset, which is numerical values.

Also, we added eight additional features to the dataset, which is based on base features. These features consist of session-based statistical features and users' campaign information related to their attribution to the game. These additional features and their characteristics are shown in Table 3.2.

Table 3.2: Additional features used in lifetime value prediction.

Feature	Description	Type	Range	Is Categorical	Missing Value Rate
Max Session Length	Maximum session duration for users in first 24 hour.	Numerical	[0,9016]	0	0%
Median Session Length	Median session duration for users in first 24 hour.	Numerical	[0,7019]	0	0%
Average Sesion Length	Average session duration for users in first 24 hour.	Numerical	[0,4812]	0	0%
Session Per Day	Session number of users per day.	Numerical	[0,1391]	0	0%
Campaign Name	The information from which campaign the user attributed	Categorical	123 different values	1	0%
Partner	Campaign Provider	Categorical	9 different values	1	0%
ECPI	Cost per install information.	Numerical	[0.01,2.38]	0	0%
Device Brand	Device brand of user's	Categorical	108 different values	0	0%

The main contribution of this study is in the Modelling section. For the Modeling section, we proposed a new ensembling process. The steps of the new model have been explained briefly below.

Step 1: The dataset is made up of objects with different characteristics. This object may be a product, a user, etc.

Step 2: The techniques of Fuzzy Clustering are applied to the data to locate related classes of each entity. Fuzzy clustering techniques provide that this object is processed according to its category knowledge, whether an item is linked to more than one category. The threshold value for membership is set in this step.

Step 3: In the third step, the entire dataset is run through candidate models with chosen parameters, and predictions are saved in data frames.

Step 4: The most efficient models (with chosen threshold output criteria) are determined for each fuzzy cluster.

Step 5: The weighted average method is used to predict each fuzzy cluster.

The formula of the proposed methodology is given below. Firstly we calculate the weights of each model, which the user chooses at the modeling as a successful model. The weight calculation is defined by Equation 3.5.

$$W(xi)= (1/e(i)) / (\sum 1/e) \quad (3.5)$$

Where e(i) denotes the RMSE of the candidate model, the calculated weight is used in the final output stage of the prediction. The output equation is defined by Equation 3.6.

$$Z= \sum w(xi)ft(xi) \quad (3.6)$$

Where w(xi) denotes the weight, ft(xi) denotes the prediction of the candidate model. This calculation step is applied to each fuzzy cluster because each fuzzy cluster's applied the most successful models and their error rates differ.

The primary reason for this approach is that there are different properties of each object in the observation dataset. Such attributes are price, amount of sales, seasonality, etc. So, a certain amount of prediction error is created by using the same model to forecast each object. To prevent this problem, a number of literature studies have been conducted. They usually apply clustering techniques that are, in particular,

K-Means clustering[30, 31]. But K-Means clustering, which is part of hard-clustering methods, assigns a single cluster to each individual. But if the object has more than one cluster attribute, this approach does not work. Fuzzy C-Means clustering, one of the most well-known fuzzy clustering methods, was thus extended to the dataset and different k initial seed values were tested for FPC[32] ratings. For a fuzzy cluster with almost a 0.9 FPC ranking, five were selected for a k number. Table 3.3 displays the characteristics and unique user counts of the clusters. Also, Table 3.3 shows us some of the users have characteristics of more than one group. So, they involve more than one group at the same time.

Table 3.3: Fuzzy cluster details.

Fuzzy Clusters	Unique User Count	Fuzzy Clusters	Unique User Count
1	156244	3,4	6285
1,2	21045	4	34200
2	310582	4,5	3415
2,3	25389	5	50402
3	47050		

Three of the most popular boosting algorithms, XGBoost, Light GBM, and CatBoost, were applied separately to the data set for regression modeling with their different hyperparameters. For each observation, the prediction outcomes were saved in separate data frames. The output results of each algorithm were then examined for each cluster. The performance results of the algorithms for each cluster are summarized in Table 3.4. The RMSE metric, a standard metric in regression problems, was used to determine prediction efficiency. According to the results in Table 4, each model with different parameters has different output effects in other clusters. As a result, combining the top three models with their optional parameters based on their reciprocal proportion of RMSE is a more feasible solution for better prediction. The ensembled prediction was then made, and its output was compared to the performance of the other model and parameter groups. Table 3.5 shows the comparison in detail.

Compared models are three of the most popular existing ensemble learning algorithms in the literature. The findings inform us that the ensembled solution, a combined version according to the fuzzy cluster distribution of existing models in the

literature, has better success at the predictive level than the model – parameter tuples individually.

This proposed methodology was firstly used in research by Tekin et al. [63] for an online travel agency's digital advertisements' click prediction with clustering hotels according to their characteristics. Fuzzy Clustering technic provided better prediction results according to applying machine learning algorithms directly to the whole dataset. Besides, Tekin and Cebi [64] compares soft clustering and hard clustering technics in further research and they proposed that using soft clustering technic like Fuzzy C-Means is successful for machine learning problems which need to cluster objects which have more than one characteristic.

Table 3.4: Model and parameters prediction results on groups (RMSE).

Model	Parameters	Grp 1	Grp 1,2	Grp 2	Grp 2,3	Grp 3	Grp 3,4	Grp 4	Grp 4,5	Grp 5
XGB	max_depth:6, learning_rate: 0.01	19.77	20.46	18.89	18.08	14.60	24.21	13.98	18.93	19.58
XGB	max_depth:6, learning_rate: 0.1	5.61	5.40	5.41	6.03	5.28	6.73	29.46	24.32	20.77
XGB	max_depth:8, learning_rate: 0.01	22.46	16.75	19.41	15.30	6.90	6.48	5.54	25.54	23.23
XGB	max_depth:8, learning_rate: 0.1	23.98	31.13	16.95	7.96	25.94	8.41	6.80	24.04	27.94
CB	depth:4, learning_rate:0.01	7.90	7.51	2.33	1.73	26.07	22.16	5.55	19.72	25.64
CB	depth:4, lr:0.03	20.25	20.08	19.70	6.40	4.93	20.69	19.31	5.31	4.84
CB	depth:8, learning_rate:0.01	20.51	20.33	21.12	14.75	5.26	5.90	5.74	6.35	25.95
CB	depth:8, learning_rate:0.03	4.12	4.86	6.56	6.70	6.20	4.47	5.62	4.18	4.93
LGBM	max_depth:6, learning_rate:0.01	5.33	4.80	5.48	4.59	4.78	20.46	15.45	4.92	5.25
LGBM	max_depth:6, learning_rate:0.1	28.55	28.95	23.00	1.88	1.67	5.68	17.55	24.03	1.85
LGBM	max_depth:8, learning_rate:0.01	1.84	2.00	1.61	6.73	7.28	31.58	31.21	23.23	22.88
LGBM	max_depth:8, learning_rate:0.1	14.66	5.62	6.46	5.94	5.93	4.64	14.15	5.15	6.46

Table 3.5: Overall prediction results with each model and parameters (RMSE).

Model	Parameters	Overall RMSE	Model	Parameters	Overall RMSE
XGB	max_depth:6, learning_rate:0.01	19.64	CB	depth:8, learning_rate:0.03	5.87
XGB	max_depth:6, learning_rate:0.1	11.04	LGBM	max_depth:6, learning_rate:0.01	7.72
XGB	max_depth:8, learning_rate:0.01	16.39	LGBM	max_depth:6, learning_rate:0.1	14.28
XGB	max_depth:8, learning_rate:0.1	19.03	LGBM	max_depth:8, learning_rate:0.01	14.69
CB	depth:4, learning_rate:0.01	13.19	LGBM	max_depth:8, learning_rate:0.1	7.70
CB	depth:4, learning_rate:0.03	13.78	Ensembled Model		4.18
CB	depth:8, learning_rate:0.01	14.08			

3.4 Conclusion

In this research, we aimed to predict customer lifetime value for a mobile game's users. This prediction is so crucial for companies for determining the marketing cost cap for companies. We proposed a new ensemble approach for the prediction that is based on fuzzy logic and multiple model selection for this purpose. Normally, before the machine learning algorithms are applied to the dataset, data preprocessing, missing value elimination and feature engineering steps are applied to the dataset. Then hyperparameter optimization is applied to the dataset for improving results. These steps are so crucial for model success. However, this method can fail for individual objects in the dataset that do not have similar characteristics. Our approach may be applied to goods or users: items, consumers, or users with various characteristics such as price, user behaviour, etc.

In our approach, we collected the gameplay and session data of users for the first 24 hours in the game in the first step. After that, we also added some additional features to the base features like users' campaign information related to their attribution to the game and some statistical session characteristics of users. In the second step, missing values are filled in the dataset with the most popular approaches in the literature like fill with zero, fill with most frequent value in the feature, etc. After that, data preprocessing and feature engineering technics are applied to the dataset like one-hot encoding for categorical variables.

In the modelling section, we clustered all users in the dataset with a fuzzy clustering technic called Fuzzy C-Means clustering. For finding the optimum value of the cluster number, the FPC score parameter was used. We split users into five groups that have the best FPC score, which is almost 0.9. For choosing fuzzy clustering instead of hard clustering technics in the literature like K-Means Clustering, some users can be related to more than one group according to their characteristics.

After the clustering process, XGBoost, Catboost and LightGBM algorithms which are so popular ensemble learning algorithms with different hyperparameters, are applied to each cluster separately. Predictions and validation performance results were stored in the data frames separately. RMSE metric, a popular evaluation metric of regression problems, was used for model evaluation.

In the last phase, these three algorithms were ensembled with their weighted average of predictions according to the performance result. These weights were determined according to their Root Mean Squared Error's reciprocal in each cluster with different hyperparameters.

The results indicate us ensembling with a fuzzy approach has better prediction performance than applying algorithms individually with their different hyperparameters. Our ensemble approach reached the minimum Root Mean Squared Error rate with a 4.18 overall Root Mean Squared Error value.

For future work, we are aiming to use our approach in different domains and different datasets. Also, to get more accurate results, we seek to use different fuzzy clustering technics in the literature instead of Fuzzy C-Means Clustering. Again, this method can be useful for classification problems which dataset consists of objects with other characteristics. Differently, this method also can be helpful in missing value prediction in the dataset. Filling missing values are also an important stage of machine learning which is crucial for model performance. Instead of filling missing values in the dataset with the most frequent value of the feature column, filling the missing value can be applied within the most frequent value in the correspondent fuzzy cluster. This can also reduce the prediction error.

4. RETENTION PREDICTION IN THE GAMING INDUSTRY: FUZZY MACHINE LEARNING APPROACH[3]

In solving a problem we encounter daily, we may not be sure of a single answer. To be sure, we ask the same question to more than one person and act according to the majority of the votes, or we weigh the votes. Ensemble learning algorithms also act in this way. They run base algorithms multiple times to solve a problem and develop a hypothesis to vote for the results. Thanks to this approach, ensemble learning algorithms have recently achieved great success in solving complex data problems. With the advancement of technology daily, a severe performance increase has been realized in the tough problem-solving times of ensemble learning algorithms.

Ensemble learning algorithms have recently been applied in both academia and private sector problems, and these studies are pretty common in the literature. Private sector problems generally consist of various fields such as production, marketing, finance, etc. Although the sector is different, the aim is to predict the unpredictable events that will occur in the future and take precautions against these situations. Therefore, establishing a successful prediction model is very crucial.

There are many methods in the literature to establish a successful model. The number of these methods is increasing day by day. Among these methods, one of the proven successful approaches is ensemble learning approaches. The primary purpose of ensemble learning algorithms is to combine multiple algorithms or algorithm results to close their weaknesses and create a more robust model. In this way, it is aimed to reduce the error rate in prediction.

Since the day it was first introduced in the literature, fuzzy logic [4] has been used in many fields such as control, optimization, and data analysis. Fuzzy logic approaches have also started to be used in machine learning over time, and these studies can be found in the literature in many different fields [76]. Fuzzy logic is generally used to

[3] This chapter is based on the paper "Retention Prediction in the Gaming Industry: Fuzzy Machine Learning Approach". Tekin, A.T. ,Çebi, F., Kaya, T., 2022. Industrial Engineering in the Age of Business Intelligence, (Ed.s) Calisir, F. Pre-press, Pre-press.

extend machine learning and data mining studies in the literature. For this purpose, fuzzy logic approaches are seen, especially in clustering and association rule mining studies.

The fuzzy machine learning approach is applied to fuzzy sets. This, of course, necessitates the extension of corresponding learning algorithms, which usually assume crisp data. Although the number of literature studies in fuzzy machine learning increases, these studies are generally carried out without clarifying the actual meaning of fuzzy observation.

The datasets used in machine learning problems that are tried to be solved also contain fuzziness. Modelling objects with different characteristics in the same data set can negatively affect prediction success. At this point, using clustering approaches and machine learning models together and modelling items with similar characteristics in similar groups can increase prediction success. As a clustering approach, fuzzy clustering approaches can also be preferred, unlike hard clustering approaches like K-Means Clustering [75].

In this study, we aim to predict the retention value of attributed users in a mobile game. Naturally, the genders, countries, age groups, and game behaviours of the users who play this game differ. These differences affect many variables, such as when the user stays in the game and how long he will play the game. In this case, clustering users is a good approach, but a user can also have the characteristics of two or more clusters at the same time. For this reason, we proposed a method that combines the fuzzy clustering method and ensemble learning method. In this way, we aimed to increase the overall prediction success in the modeling phase compared to using ensemble learning methods alone. In addition, since each algorithm - parameter group can have different success rates on different clusters, we aimed to make the prediction more successful by weighting the most successful algorithm - parameter groups with the prediction success rates for each cluster.

The paper addresses the literature review of ensemble learning and fuzzy methods of machine learning in Section 4.2. Our proposed approach which ensembles fuzzy clustering and ensemble learning algorithms and modelling data are illustrated in Section 4.3. Finally, the study's results are briefly outlined, and the last segment presents potential work.

4.1 Literature Review

4.1.1 Ensemble learning

In machine learning problems, a single-week learning algorithm is very dependent on the training set, which can lead to overfitting [80]. It aims to close these weaknesses by aggregating more than one weak algorithm to deal with this problem. This method is called ensemble learning [79]. Ensemble learning approaches, whose importance and use have increased considerably in recent years, are used in many studies in the literature. They have been very successful in real-life applications in machine learning and machine learning challenges such as Kaggle competitions [67, 74, 78]. One of the most important reasons for the ensemble learning approach's success is the generalization ability [73]. Ensemble learning methods are divided into two groups as parallel and sequential methods. Weak learners are generated in parallel with the bagging approach in the parallel ensemble learning method. In the sequential ensemble learning method, weak learners are generated sequentially with the boosting approach [79]. Bagging and boosting approach principles are introduced briefly below.

4.1.1.1 Bagging

Bagging, which stands for bootstrap aggregating, is a technique that trains various homogeneous weak learners separately from each other in parallel and then combines them using a deterministic averaging procedure to achieve the final prediction or classification [70]. In this method, while creating regression trees, a training set of the same size as the original data set is created based on the "bootstrap" approach. Some items can be left out in this training set, while others can also be used repeatedly. Breiman stated that for the created bags to be effective, the observation data in the bag is unstable; that is, it depends on the rate of response to changes in the training data [35]. Random Forest is an example of bagging type ensemble learning algorithms.

Random forest

Random Forest (RF) comprises many individual decision trees that work together to form an ensemble. RF is an ensemble of Classification and Regression Trees (CART) [68] trained on datasets of the same size as the training set, known as

bootstraps generated by random resampling on the training set. In RF, the prediction is based on each tree's predictions. So, the most votes become the prediction. The random forest philosophy is that many autonomous decision tree models working as a group outperform any single decision tree model. After constructing a tree, a set of bootstraps do not contain any specific record from the original dataset. Because out-of-bag samples are used as the test dataset. RF model can be represented as

$$\hat{m}(x) = \frac{1}{M}\sum_j \hat{m}\, j(x) \tag{4.1}$$

where $\hat{m}j$ denotes an individual tree, and the prediction is based on the averaging of each tree's prediction.

4.1.1.2 Boosting

Boosting, which stands for sequential ensemble learning approach, can be used for both classification and regression. It produces a weak prediction model at each stage, which is then weighted and applied to the overall model, reducing variance and bias and improving model efficiency. Adaptive Boosting (Adaboost), Gradient Boosting, Extreme Gradient Boosting (XGBoost), Catboost and Light Gradient Boosting Machine (Light GBM) are the most popular boosting type ensemble learning algorithms. These algorithms are explained briefly below.

Adaboost

Adaboost is an ensemble-type boosting algorithm that Freund and Shapire proposed in 1995 [71]. Adaboost is an iterative algorithm, and it generates a robust model from a set of weak models. In each generation, it tries to minimize the sum of the training error. Adaboost model can be represented as

$$E_t = \sum_i E[F_{t-1}(x_i) + a_t h(x_i)] \tag{4.2}$$

$F_{t-1}(x)$ is the previous learner, E is an error function, and $a_t h(x)$ is the weak learner, contributing to the stronger learner [72].

LightGBM

The Light Gradient Boosting algorithm is a new version of the GBDT algorithm. It is widely used in a wide range of modelling problems, including classification and regression. LightGBM employs two new strategies to accommodate many data instances and functions: gradient-based one-side sampling and exclusive function

bundling [66]. Compared to base gradient boosting strategies or Extreme Gradient Boosting, LightGBM extends the decision tree vertically, whereas others extend it horizontally. This feature enhances LightGBM's ability to process large amounts of data.

Catboost

Catboost is a new proposed version of the gradient boosting type algorithm, which Prokhorenkova proposes in 2018 [60]. Catboost works reliably with categorical features with the least amount of information loss. CatBoost is distinct from other gradient boosting algorithms such as Extreme Gradient Boosting and LightGBM. It employs ordered boosting, an effective modification of gradient boosting algorithms, to address target leakage [65]. Furthermore, Dorogush claims that Catboost attempts to avoid overfitting issues created by gradient boosting algorithms by executing random permutations to change leaf values when deciding on a tree structure. Catboost model can be represented as

$$Z = H(x_i) = \sum J_j = 1c_j 1\{x \in R_j\} \tag{4.3}$$

$H(x_i)$ is a decision tree function of the explanatory variables x_i, and R_j is the disjoint region corresponding to the tree leaves [60].

XGBoost

XGBoost is a well-known gradient boosting algorithm suggested by Chen and Guestrin [10]. XGBoost is an improved GBDT algorithm that utilizes many decision trees and is commonly used in classification and regression. XGBoost incorporates a regularization principle to optimize the size of the tree's classification function in order to make it more reproducible. Furthermore, regularization aids in the prediction of feature value, which is critical in big data problems. XGBoost can be represented as

$$Z = F(x_i) = \sum T_t = 1f_t(x_i) \tag{4.4}$$

where x_i denotes the explanatory variables and $f_t(x_i)$ is the output function of each tree.

Gradient boosting

Gradient Boosting is the base version of XGBoost, Catboost, and LightGBM, and it is proposed by Friedman [13]. It is a tree-based machine learning algorithm that

applies the gradient boosting method effectively and scalably. It is also one of the most common machine learning algorithms, and it has been stated in the literature that it performs very well in both regression and classification problems. Gradient Boosting attempts to minimize bias variation by using the base learner as a weighted total and reweighting misclassified results. Besides, it employs decision trees as base learners to reduce the loss function. Gradient Boosting can be represented as

$$Z = F(x_i) = \sum M_j = 1\beta_j\, h \tag{4.5}$$

where the function $h(x;\, b_j)$ is the base learner, x is the explanatory variables, β_j is the expansion coefficients, and b_j is the parameters of the model.

4.1.2 Evaluation of performance metrics

In classification-type machine learning problems, we apply several machine learning algorithms for choosing the best appropriate model. For algorithm performance comparison, we consider some metrics such as accuracy, precision, AUC, F1 score, etc. These metrics are explained briefly below.

Accuracy: It is a metric that is one of the most popular metrics for evaluating classification-type machine learning problems. This score indicates the algorithm's performance by showing the true value ratio of the predicted label. So, it shows the algorithm's overall performance.

Precision: It is a metric that shows us prediction's power. Precision is defined as the proportion of correctly predicted positive observations to all predicted positive observations.

Recall: Recall is named as sensitivity also. It is the ratio of correctly predicted positive observations to all observations in the actual class. It shows us the effectiveness of the algorithm in a single class.

AUC: AUC stands for Are Under the ROC Curve. ROC (Receiver Operation Characteristics) is a curve that shows the performance of a classification model. AUC is a metric that aggregates output overall classification thresholds. AUC can be translated as the model's chance to score a random positive example higher than a random negative example.

F1 Score: It is a metric that the weighted average of Precision and Recall. F1 Score is considered both false positives and negatives. F1 score can be more useful, especially in imbalanced datasets [77].

4.1.3 Fuzzy clustering techniques

Fuzzy logic, an essential place in the literature, is often widely used in machine learning problems. One of the primary reasons for this is that the answers to real-life problems do not consist of only 0 and 1 singular values, as is the case with machine learning approaches that address real-world problems.

Fuzzy Logic [4, 51, 52] was discovered by Lofti Zadeh in 1965. Zadeh believed that all real-world problems could be solved using analytical and computer-based methods [53]. In 1964, he revealed the "Fuzzy Set Theory." Some academic communities have criticized this theory for its uncertainty, but it is used in many different areas today. A fuzzy set is defined by a function that maps objects in the respective domain to the membership value in the set [53]. Basically, Fuzzy Logic is a method that allows defining intermediate values between traditional evaluations such as true/false or yes/no.

The fuzzy logic theorem is also frequently encountered in the literature with clustering problems, one of the unsupervised learning problems. Clustering algorithms can generally be considered in two classes as hard and soft clustering [56]. In the hard clustering method, each observation in the test data set belongs to only one cluster, while in the soft clustering method, an observation may belong to more than one cluster [50]. The soft clustering method is also included in the literature as fuzzy clustering. In the fuzzy clustering method, the membership level is calculated for each observation value. This level lies between 0 and 1.

There are different approaches for fuzzy clustering suggested in the literature. These approaches are Fuzzy C-Means Clustering (FCM), Possibilistic C-Means Clustering (PCM), Fuzzy Possibilistic C-Means Clustering (FPCM), Possibilistic Fuzzy C-Means Clustering (PFCM). These algorithms are explained briefly below.

Fuzzy C-Means Clustering (FCM): FCM is a clustering technique in which each data point belongs to a cluster to a degree determined by a membership degree. Jim Bezdek proposed this approach in 1981 [28].This approach is an advancement to previous clustering approaches. FCM groups data points as populating

multidimensional space to a set number of different clusters. The primary advantage of FCM is that it allows data points for memberships of each cluster with degrees between 0 and 1.

FCM is based on a minimization objective function. This function can be represented as

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 , 1 \leq m < \infty \quad (4.6)$$

where u_{ij} is the degree of membership of x_i in the cluster, x_i denotes that i th of d-dimensional measured data, c_j is the dimension center of the cluster [69].

Possibilistic C-Means Clustering (PCM): This method, first proposed by Krishnaparum and Keller [57]. in 1996, enables the detection of outliers in the data set [50]. In this method, low typicality values are produced for outliers, and these outliers are automatically eliminated. This method is also sensitive to initial assignments and can assign an observation value to more than one cluster at the same time because it exhibits a flexible clustering approach. Furthermore, typicalities can be extremely sensitive to selecting the additional parameters required by the PCM model.

Fuzzy Possibilistic C-Means Clustering (FPCM): This method, proposed by Pal, Keller, and Bezdek [58]. in 1997, consists of combining the FCM and PCM approaches. This method has been tried to obtain more successful results y combining the typical values of the FCM clustering method with the typical values of the PCM clustering method. FPCM normalizes the possibility values such that the number of all data points in a cluster equals 1. Although FPCM is much less susceptible to the problems that FCM and PCM cause, the probability values become very small as the data set size increases.

Possibilistic Fuzzy C-Means Clustering (PFCM): This method was suggested by Pal et al. [59] in 2005; They aimed to exhibit a more successful clustering approach by eliminating the lack of noise sensitivity of the FCM algorithm, the random clustering problem of the PCM approach, and the row sum constraints of the FPCM clustering approach. Pal et al. derive the first-order necessary conditions for PFCM objective function extrema and use them as the foundation for a typical alternating optimization approach to finding PFCM objective functional local minima.

Soft and hard clustering methods are used quite frequently in machine learning problems, and although they are used on their own, they are also used as a precursor in regression or classification problems. When the points in the data set are grouped according to their behavioural characteristics, making individual model trials for these groups can increase the success of the general model.

4.1.4 Proposed methodology

Churn prediction is one of the most popular topics in classification problems. Instead of churned customers, companies have a significant focus on their retained users. More than half of the revenue, especially in the game industry, comes from users who continue to play after the day they downloaded the game. While the revenues from the users who churn from the first day cannot cover their costs, retained users have an essential place in the established revenue model.

The dataset used in this study is related to a crossword puzzle game published in Google Play Store and App Store. This dataset consists of users' first 24 hours of gameplay data and session information. In this study, we aim to predict customers are retained or not after the first 24 hours. Base features that are used in this study are shown in Table 4.2. The dataset consists of 24 features and 356603 rows, indicating each customer's first 24-hour activity summary.

Firstly, data cleaning, data preprocessing, missing value elimination were administered to the dataset. After that, one hot encoding technique was applied to the dataset for the categorical values, and all categorical values were converted to numerical values. Then, we used the min-max scaling technique for the numerical values, and all of the data were scaled. Traditional machine learning algorithms and boosting type ensemble learning algorithms were applied to the dataset with default parameters to detect the most successful algorithms. The results can be shown in Table 4.1. The results indicate that boosting type ensemble learning algorithms are more successful than traditional algorithms, and XGBoost is the most accurate prediction algorithm.

Table 4.1: Algorithm performances with default parameters.

Model	Accuracy	AUC	Recall	Prec.	F1
XGB	0,828	0,913	0,709	0,881	0,786
LGBM	0,826	0,910	0,711	0,877	0,785
CB	0,818	0,897	0,755	0,835	0,793
GBC	0,803	0,884	0,707	0,835	0,766
Adaboost	0,779	0,856	0,723	0,787	0,754

The main contribution of this study is in the modelling section. In the modelling phase, instead of directly applying traditional machine learning algorithms to the final data we have, firstly, similar users were grouped with the fuzzy clustering approach. The Fuzzy C-Means Clustering method was preferred instead of the K-Means clustering approach. The degree of belonging of a user to each cluster was calculated separately, and users above a certain value were included in more than one cluster at the same time. In this way, after the clustering approach, the dimensions of the clusters shrank less than the K-Means clustering approach, but still, the users were grouped with similar users.

Table 4.2: Base features used in retention prediction.

Feature	Description	Type	Range	Missing Value Ratio(%)
session_cnt	The number of sessions started for the user in the first 24 hours.	Numerical	[1,82]	0,00
session_length	The total duration of sessions started for the user in the first 24 hours.	Numerical	[0,17540]	0,00
app_version	Application version of the game	Categorical	13 different values	0,00
language	The language which users played in the first 24 hour	Categorical	30 different values	0,00
max_level_no	The maximum level number which the user reached	Numerical	[1,459]	0,00
gameplay_duration	The total duration of levels for the user in the first 24 hour	Numerical	[1,14760]	0,00
bonus_cnt	The number of bonuses which the user used	Numerical	[0,109]	0,00
hint_cnt	The number of hints which the user used	Numerical	[0,124]	0,00
repeat_cnt	The number of repeating the levels which users used for completing levels	Numerical	[0,1086]	0,00
gold_cnt	The final gold amount which the user has at the end of the first 24 hour	Numerical	[0,525461]	0,00
banner_cnt	The number of banner-type advertisements that users display in the first 24 hours.	Numerical	[0,1679]	0,00
interstitial_cnt	The number of interstitial type advertisements which users display in the first 24 hours.	Numerical	[0,444]	0,00
rewarded_video_cnt	The number of rewarded video type advertisements that users display in the first 24 hours.	Numerical	[0,25]	0,00
revenue	The revenue amount is acquired from users in the first 24 hours.	Numerical	[0,10.82]	0,00
max_session_length	Maximum session duration for users in the first 24 hours.	Numerical	[0,6060]	0,00
avg_session_length	Median session duration for users in the first 24 hours.	Numerical	[0,4149]	0,00
campaign_name	The information from which campaign the user attributed	Categorical	158 distinct values	2,33
partner	Campaign provider information	Categorical	5 different values	2,33
ecpi	The acquisition cost of the user	Numerical	[0,2.02]	0,00
os	The operating system information of the user	Categorical	2 different values	0,00
country	The country information of the user	Categorical	149 different values	0,00
retention	Is the customer retained or not?	Label	2 different values	0,00

In the Fuzzy Clustering part, different k initial seed values were tested for FPC [32] ratings. Also, different "m" fuzzifier parameters 1.2, 1.5, 2, 2.5 were tested to find optimum cluster structure. For a fuzzy cluster with a 0.85 FPC ranking, c=3 and m=2 were selected as the initial seed value and fuzzifier parameters. Table 4.3 shows us some of the users have characteristics of more than one group. So, they involve more than one group at the same time.

Table 4.3: Fuzzy cluster details.

Cluster a	Cluster a-b	Cluster a-c	Cluster b	Cluster b-c	Cluster c
110286	43601	31574	76957	35833	58352

After that, the three most successful algorithms, XGBoost, Catboost, and LightGBM were applied to each fuzzy cluster with their different hyperparameters. The results show us each algorithm with different parameters has different success ratios in each cluster. So, we can say that there is no best algorithm for the entire dataset. For each observation, the prediction outcomes were saved in separate data frames. The output results of each algorithm were then examined for each cluster. The performance results of the algorithms for each cluster are summarized in Table 4.4.

Table 4.4: Algorithm performances with different parameters.

Model	Parameters	Cluster a	Cluster a-b	Cluster a-c	Cluster b	Cluster b-c	Cluster c
XGB	md:6,lr:0.3	0,81	0,79	0,78	0,8	**0,82**	**0,83**
XGB	md:6,lr:0.1	0,8	0,8	**0,83**	0,77	0,77	0,81
XGB	md:8,lr:0.3	0,82	0,79	0,79	0,76	0,81	0,8
XGB	md:8,lr:0.1	0,8	0,78	0,8	0,82	**0,82**	**0,83**
CB	d:4,lr:0.01	0,79	**0,82**	0,8	0,76	0,81	0,81
CB	d:8,lr:0.01	0,79	0,81	0,76	0,78	0,81	0,79
CB	d:4,lr:0.1	0,78	0,8	0,77	0,8	0,78	0,79
CB	d:8,lr:0.1	0,78	0,78	0,81	0,79	0,8	0,78
LGBM	lr:0.1,ne:200	0,82	0,81	0,77	0,79	0,76	0,81
LGBM	lr:0.1:ne:100	0,81	0,79	0,79	0,81	**0,82**	0,8
LGBM	lr:0.3,ne:200	**0,83**	**0,82**	0,8	**0,83**	0,81	0,79
LGBM	lr:0.3:ne:100	0,82	**0,82**	**0,83**	0,81	0,79	0,81

In our proposed method, we chose accuracy as a performance metric that is popular in classification problems. The steps of our proposed method have been explained briefly below.

Step 1: Data preprocessing and feature engineering technics that are necessary for modelling are applied to the dataset.

Step 2: FCM is applied to the processed dataset, and a membership value threshold is set to cluster objects. According to this threshold value, objects can be assigned to more than one cluster.

Step 3: Candidate models with different parameters are applied to all clusters separately, and the best models are chosen according to the selected success criteria for each cluster. At this stage, 70% of the dataset is used as learning and 30% as a test set. The success rate is based on the accuracy rate of the test set.

Step 4: The weighted average method is used to ensemble the outcomes of the best models. The weight calculation is defined by Equation 4.7.

$$\text{Weight}_{c_i} = \frac{\text{acc}_{c_i}}{\sum_{j=1}^{n} \text{acc}_{c_j}} \tag{4.7}$$

Where c_i denotes each classifier and acc_{c_i} is the accuracy of c_i, j is the number of classifiers in the ensemble group and acc_{c_j} denotes the j^{th}classifier.

As a result, combining the top three models with their optional parameters based on their accuracy is a more feasible solution for better prediction. The ensembled prediction was then made, and its output was compared to the performance of the other model and parameter groups. Table 4.5 shows the comparison in detail.

Table 4.5: Overall prediction results with each model and parameters (accuracy).

Model	Parameters	Overall Accuracy	Model	Parameters	Overall Accuracy
XGB	md:6,lr:0.3	0,81	CB	d:8,lr:0.1	0,80
XGB	md:6,lr:0.1	0,80	LGBM	lr:0.1,ne:200	0,80
XGB	md:8,lr:0.3	0,80	LGBM	lr:0.1:ne:100	0,81
XGB	md:8,lr:0.1	0,81	LGBM	lr:0.3,ne:200	0,82
CB	d:4,lr:0.01	0,80	LGBM	lr:0.3:ne:100	0,82
CB	d:8,lr:0.01	0,79	Ensembled Model		0,85
CB	d:4,lr:0.1	0,79			

The findings show us our proposed ensemble solution, a combined version according to the fuzzy cluster distribution of existing models in the literature, has better success at the predictive level than the model–parameter tuples individually.

4.1.5 Conclusion

This study aims to predict whether mobile game users will stay in the game after the first 24 hours. This prediction affects the user-based marketing strategies of gaming

companies in the upcoming period. It is of great importance in the investment processes of these companies, which have received severe investments recently.

For this prediction, we proposed a new method combining fuzzy Logic and community learning steps. This method aims to identify one or more clusters to which they belong with fuzzy clustering logic rather than evaluate the users within a single cluster and model them separately. It is aimed to establish a more successful model by combining the results of the most successful models among the models created independently, in direct proportion to their accuracy.

In this study, the game movements of the users of a mobile game company in the first 24 hours and the data of the advertisements they see while downloading the game to their phones are used. First, the missing data in the available data set was filled, and then the categorical data for the modelling stage were digitized. In the next step, the entirely digitized data were scaled.

In the modelling phase, the data set we have was modelled with the default parameters of various algorithms, and the results were examined. In the next step, the first step of the proposed method, fuzzy clustering, was applied, and the data was separated into three main clusters and sub-components of these clusters. Thanks to this stage, users who could belong to more than one group simultaneously were determined, and these users were evaluated in separate clusters.

After the clustering process, XGBoost, Catboost, and LightGBM, the three most successful models in the first modelling process, were applied separately to the groups with various hyperparameter combinations. It was observed that different groups were different models in each group that could be more successful.

The estimates of each group's three most successful model combinations were combined with the weighted average method at the last stage. The weights were determined by the accuracy rates of the models. When the ensemble model’s success rates are compared with the success achieved when the models are used one by one, an increase in performance was observed in the results.

In future work, it is aimed to apply this proposed method in different fields and different datasets. Besides, other fuzzy clustering approaches suggested in the literature will be added to the algorithm-cluster combinations in the modelling clustering stage to achieve more successful results.

5. CONCLUSIONS

Companies manage marketing activities with large budgets for profit in the private sector. Today, data analytics studies and machine learning modelling have become inevitable for more effective campaign management, which entails many financial costs for companies. In this way, companies can make more profitable process management with a more effective budget and campaign management. For this reason, companies need to make predictions in this area with less error rate. For this purpose, ensemble learning models, which are essential in the literature, increase their importance day by day.

On the other hand, the fuzzy clustering approach has been applied in many fields for many years and has been accepted its success in the literature. The advantages and disadvantages of fuzzy clustering methods are compared with hard clustering methods. In this study, ensemble learning methods were applied by combining the fuzzy clustering approach to datasets consisting of items similar to each other as characteristic features. It is aimed to increase the success rate of the models according to the individual application of the algorithms.

In the second chapter, an online travel agency's sales and cost prediction are realized. In this direction, firstly, data preprocessing and feature engineering steps are applied to the raw data set. As a result of these processes, ensemble learning algorithms were applied to the dataset. The results were compared in Table 2.4 and Table 2.6. Then, the clustering process was applied to the dataset because the hotels in the dataset are similar to or different from other hotels in terms of characteristic features. The K-Means clustering method, one of the hard clustering methods, and the Fuzzy C-Means clustering method, which is one of the fuzzy clustering methods, were applied to the dataset. The modelling process was repeated by trying their combinations with the three most successful algorithms in the modelling phase. The fuzzy C-Means clustering method produced more successful results than both the model created using the K-Means clustering method and the models applied directly to the dataset. The final results obtained are compared in Table 2.10.

In the third chapter, within the scope of the study applied in the second chapter, it was decided to apply this method for a different problem in another sector, since combining the Fuzzy C-Means clustering method with the ensemble learning techniques provides an increase in performance in the modelling phase. In this context, it has been tried to predict the lifetime value of the players by using the user behaviour and marketing data of a mobile game. According to the first study, the working principles of ensemble learning methods were examined in more depth, and some algorithms that were not used in the first study were included in the study. At the same time, hyperparameter optimization, which has an essential place in machine learning problems, was applied separately to fuzzy set-algorithm groups. The results were compared in this way. In this study, the proposed model by combining ensemble learning algorithms and fuzzy clustering method produced more effective outcomes than applying the algorithms individually to the entire dataset. The results are compared in Table 3.5.

In the fourth chapter, unlike the regression problems in the second and third chapters, the proposed method is tested in a classification problem. For this purpose, it has been tried to predict whether users will continue to play the game at the end of the first 24 hours over a different dataset of the mobile game in the second study. In this study, in addition to the hyperparameter optimization applied to increase the performance of the models, different fuzziness parameters were also tested in the fuzzy clustering stage for a more efficient clustering process, and fuzzy clustering was performed with the combination with the most optimum FPC score. In combining the model members, a weighting approach that is directly proportional to the accuracy rate was preferred instead of the weighting approach inversely proportional to the error rate in the regression. As a result, the proposed model produced more effective outcomes than applying the algorithms individually to the entire dataset in this study. The results are compared in Table 4.5.

For the discussion, the proposed method divides the data into clusters and model them separately. So, the total workload and the total execution time in the modelling is higher than modelling all data alone. On the other hand, the improvement supplied by the proposed method can be crucial for the companies. Thanks to the improvement gained in the modelling phase, companies that will use these predictions as strategic planning can gain significant advantages and profits. Another

issue is that studies to be carried out on different datasets with different characteristics should be increased to test the validity of the proposed model. Also, for the effective usage of the proposed model, it is more appropriate for the data set used in the model to consist of features with more than one characteristic.

We suggest that this proposed method be applied to different datasets from different sectors for further research. Still, new studies with other clustering methods indicated in the literature, which are PCM [57], FPCM [58], PFCM [59] at the fuzzy clustering stage, can be carried out. Also, in further research, Fuzzy K-Means clustering [99] can be applied to the dataset instead of Fuzzy C-Means Clustering. On the other hand, the DBScan algorithm, etc., can be used as an alternative to hard clustering techniques. So, in this way, it will be possible to make a more detailed and safer result comparison in the clustering stage.

For the model validation, to achieve more reliable results at the point of comparison of the results after the modelling stage, statistical tests can be applied to the model outputs and the analyzes can be further deepened. At the same time, it will be possible to carry out modelling studies with a higher success rate thanks to this approach, which can also be used with new machine learning algorithms that can be used soon with rapidly developing technology.

REFERENCES

[1] **Rokach, L.** (2005). Ensemble Methods for Classifiers. In O. Maimon, L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 957-980). Springer.

[2] **Dimitriadou, E., Weingessel, A., & Hornik, K.** (2003). A Cluster Ensembles Framework. In *Design and Application of Hybrid Intelligent Systems* (pp. 528-534). Amsterdam, Netherlands: IOS Press.

[3] **Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., & Amancio, D. R.** (2019). Clustering algorithms: A comparative approach. *PLOS ONE 14* (1), e0210236.

[4] **Zadeh, L.** (1965). Fuzzy sets. *Information and Control, 8* (3), 338-353.

[5] **Tekin, A., & Çebi F.** (2019). Click and Sales Prediction for Digital Advertisements: Real-World Application for OTAs. In C. Kahraman, S. Cebi, S. Cevik Onar, B. Oztaysi, A.Tolga, I. Sari (Eds.), *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making. INFUS 2019. Advances in Intelligent Systems and Computing* (Vol. 1029). Cham: Springer.

[6] **Adam-Bourdarios, C., Cowan, G., Germain-Renaud, C., Guyon, I., K´egl, B., & Rousseau, D.** (2015). The Higgs Machine Learning Challenge. In *Proceedings of the NIPS 2014 Workshop on High-energy Physics and Machine Learning, in PMLR, 42*, 19-55

[7] **Babajide Mustapha, I., & Saeed, F.** (2016). Bioactive molecule prediction using extreme gradient boosting. *Molecules, 21* (8), 983.

[8] **Balfer, J., & Bajorath, J.** (2015). Systematic artifacts in support vector regression-based compound potency prediction revealed by statistical and activity landscape analysis. *PloS One, 10* (3), e0119301.

[9] **Breiman, L.** (2001). Random forests. *Machine Learning, 45* (1), 5-32.

[10] **Chen, T., & Guestrin, C.** (2016). Xgboost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining* (pp. 785-794). ACM.

[11] **Chen, T., He, T., & Benesty, M.** (2015). *Xgboost: Extreme Gradient Boosting*. R package version 0.4-2, p. 1-4.

[12] **Cootes, T. F., Ionita, M. C., Lindner, C., & Sauer, P.** (2012). Robust and accurate shape model fitting using random forest regression voting. *European Conference on Computer Vision* (pp. 278-291). Springer.

[13] **Friedman, J. H.** (2001). Greedy function approximation: A Gradient boosting machine. *Annals of Statistics*, 1189-1232.

[14] **Geurts, P., Ernst, D., & Wehenkel, L.** (2006). Extremely randomized trees. *Machine Learning, 63* (1), 3-42.

[15] **Hengl, T., Mendes de Jesus, J., Heuvelink, G. B., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., ... & Kempen, B.** (2017). Soilgrids250m: Global gridded soil information based on machine learning. *PLoS One, 12* (2), e0169748.

[16] **Malani, J., Sinha, N., Prasad, N., & Lokesh, V.** (n.d.). *Forecasting bike sharing demand*. Retrieved from https://www.semanticscholar.org/paper/Forecasting-Bike-Sharing-Demand-Malani-Sinha/1b67300836eb4ee3532f41bab6b9ed5a77676f87

[17] **Mangal, A., & Kumar, N.** (2016). Using big data to enhance the bosch production line performance: A kaggle challenge. *Big Data (Big Data), 2016 IEEE International Conference* (pp. 2029–2035). IEEE.

[18] **Nabi-Abdolyousefi, R.** (2015). *Conversion rate prediction in search engine marketing.* (Doctoral dissertation).

[19] **Richardson, M., Dominowska, E., & Ragno, R.** (2007). Predicting Clicks: Estimating the Click-Through Rate for New Ads.In *Proceedings of the 16th International Conference on World Wide Web, WWW '07* (pp. 521–530). New York, NY, USA: ACM.

[20] **Ridgeway, G., Madigan, D., & Richardson, T.** (1999). Boosting methodology for regression problems. *AISTATS*.

[21] **Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P.** (2003). Random forest: A classification and regression tool for compound classification and qsar modeling. *Journal of Chemical Information and Computer Sciences, 43* (6),1947–1958.

[22] **Wang, F., Suphamitmongkol, W., & Wang, B.** (2013). Advertisement click-through rate prediction using multiple criteria linear programming regression model. *Procedia Computer Science, 17*, 803–811.

[23] **Zhou, Z. H., & Feng, J.** (2017). Deep Forest: Towards an Alternative to Deep Neural Networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence* (pp. 3553-3559).

[24] **Jain, A.K.** (2010). Data clustering: 50 years beyond K-means. *Pattern Recogn Lett*, *31*(8), 651–666.

[25] **Cai, Z., Heydari, M., & Lin, G**. (2005). Clustering binary oligonucleotide fingerprint vectors for DNA clone classification analysis. *J Comb Optim*, *9*(2), 199–211.

[26] **Cai, Z., Xu, L., & Shi, Y**. (2006). Using Gene Clustering to Identify Discriminatory Genes With Higher Classification Accuracy. In *Proceedings of the 6th IEEE Symposium on Bioinformatics and Bioengineering* (pp.235–242), Arlington, VA.

[27] **Cai, Z., Goebel, R., & Salavatipour, M. R.** (2017). Selecting Genes With Dissimilar Discrimination Strength for Sample Class Prediction. In *Proceedings of the Asia-Pacific Bioinformatics Conference* (pp.81–90).

[28] **Bezdek, J.** (1981). *Pattern Recognition With Fuzzy Objective Function Algorithms.* New York: Plenum Press.

[29] **Dunn, J. C.** (1974). A fuzzy relative of the ISODATA process and its use in detecting compact, well-separated clusters, *J. Cybernetics, 3,* 32-57

[30] **Pollard, K. S., & Van Der Laan, M. J.** (2002). A method to identify significant clusters in gene expression data. In U.*C. Berkeley Division of Biostatistics Working Paper Series* (p. 107).

[31] **Kaufman, L., & Rousseeuw, P.** (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: J. Wiley & Son.

[32] **Xie, X. L., & Beni, G.** (1991). A validity measure for fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13* (8), 841-847.

[33] **Ye, Y., Gao, J., & Shao, Y.** (2020). Robust support vector regression with generic quadratic nonconvex ε-insensitive loss. *Journal of Applied Mathematical Modelling, 82*, 235-251.

[34] **Jiang, H., Zheng, W., & Luo, L**. (2019). A two-stage minimax concave penalty based method in pruned AdaBoost ensemble. *Journal of Applied Soft Computing, 83,* Article 105764.

[35] **Breiman, L.** (1996). Bagging predictors. *Journal of Mach Learn.. Volume 40,* 24-123.

[36] **Li, Y., Zou, C., Berecibar, M., & Nanini-Mauri, E**. (2018). Random forest regression for online capacity estimation of lithium-ion batteries. *Journal of Applied Energy, 232*, 197-210.

[37] **Song, K., Yan, F., & Ding, T.** (2020). A steel property optimization model based on the XGBoost algorithm and improved PSO. *Journal of Computational Materials Science, 174,* Article 109472.

[38] **Zhou, Z. H.** (2009). Ensemble Learning. In Li S.Z., Jain A. (Eds.), *Encyclopedia of Biometrics*. Boston, MA: Springer.

[39] **Che, D., Liu, Q., Rasheed, K., & Tao, X.** (2011). Decision Tree and Ensemble Learning Algorithms With Their Applications in Bioinformatics. In H.R. Arabnia & Q.N. Tran, (Eds.), *Software Tools and Algorithms for Biological Systems* (vol. 696, pp. 191-199). Springer.

[40] **Yu, H., & Ni, J.** (2014). An improved ensemble learning method for classifying high-dimensional and imbalanced biomedicine data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics, 11,* (4), 657-666.

[41] **Mendes-Moreira, J., Soares, C., Jorge, A. M., & de Sousa, J. F**. (2012). Ensemble approaches for regression: A survey. *ACM Comput. Surv. 45,* 1, Article 10. doi:10.1145/2379776.2379786

[42] **Friedman, J. H.** (1999). Stochastic gradient boosting. *Computational Statistics and Data Analysis, 38*, 367–378

[43] **Webb, G. I. (2000).** Multiboosting: A technique for combining boosting and wagging. *Machine Learning, 40 (2).*

[44] **Freund, Y., & Schapire, R.** (1996). Experiments With a New Boosting Algorithm. Machine Learning. In *Proceedings of the 13th International Conference* (pp. 148–156).

[45] **Moisen, G. G., Freeman, E. A., Blackard, J. A., Frescino, T. S., Zimmermann, N. E., & Edwards, T. C.** (2006). Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological Modelling, 199* (2), 176–187.

[46] **Bauer, E., & Kohavi, R.** (1999). An empirical comparison of voting classification algorithms: Bagging, boosting and variants. *Machine Learning*.

[47] **Yildirim, P., Birant, K. U., Radevski, V., Kut, A., & Birant, D**. (2018). Comparative analysis of ensemble learning methods for signal classification. *26th Signal Processing and Communications Applications Conference (SIU).*

[48] **Džeroski, S., & Ženko, B.** (2004). Is combining classifiers with stacking better than selecting the best one?. *Machine Learning 54*, 255–273

[49] **Bulut, F.** (2017). Örnek tabanlı sınıflandırıcı topluluklarıyla yeni bir klinik karar destek sistemi, *Journal of the Faculty of Engineering & Architecture of Gazi University, 32* (1), 65-76.

[50] **Sujamol, S., Ashok, S., & Kumar, U. K.** (2017). Fuzzy based machine learning: A promising approach, CSI communications, *Knowledge Digest for IT Community, 41*, 8, 21-25.

[51] **Zadeh, L. A.** (1973). Outline of a new approach to the analysis of complex systems and decision processes. *Man, and Cybernetics, 1*, 28-44.

[52] **Zadeh, L. A.** (1968). Fuzzy algorithms. *Info. & Ctl., 12*, 94-102.

[53] **Kumar, M., Misra, L., & Shekhar, G.** (2015). A survey in fuzzy logic: An introduction. *IJSRD - International Journal for Scientific Research & Development, 3*, 6.

[54] **Holeček, P., Talasová, J., & Stoklasa, J.** (2011). Fuzzy Classification Systems and Their Applications. In *Proceedings of the 29th International Conference on Mathematical Methods in Economics 2011* (pp. 266 – 271). At Janská Dolina, Slovakia.

[55] **Kuncheva, L. I.** (2000). *Fuzzy Classifier Design*. Heidelberg: Springer-Verlag.

[56] **Nasibov, E., & Ordin, B.** (2019). An incremental fuzzy algorithm for data clustering problems. *Journal of Balıkesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 21*, 169-183

[57] **Krishnapuram, R., & Keller, J. M.** (1996). The possibilistic C-means algorithm: insights and recommendations. *IEEE Transactions on Fuzzy Systems, 4* (3), 385-393.

[58] **Pal, N. R., Pal, K., & Bezdek, J. C.** (1997). A mixed c-means clustering model. In *IEEE International Conference Fuzzy Systems* (pp. 11 -21).

[59] **Pal, N. R., Pal, K., Keller, J. M., & Bezdek, J. C.** (2005). A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems, 13* (4), 517–530. doi:10.1109/tfuzz.2004.840099

[60] **Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A.** (2018). Catboost: Unbiased boosting with categorical features. *32nd Conference on Neural Information Processing Systems*. Montreal, Canada.

[61] **LightGBM**. (2017). *Optimal split for categorical features*. Retrieved from http://lightgbm.readthedocs.io/en/latest/Features.html#optimal-split-for-categorical-features.

[62] **LightGBM**. (2017). *Categorical feature support*. Retrieved from http://lightgbm.readthedocs.io/en/latest/Advanced-Topics.html#categorical-feature-support.

[63] **Tekin, A.T., Kaya, T., & Çebi, F.** (2020) Click Prediction in Digital Advertisements: A Fuzzy Approach to Model Selection. In: Kahraman, C., Cevik Onar, S., Oztaysi, B., Sari, I., Cebi, S., Tolga, A., (Eds.), *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions* (Vol 1197). Springer, Cham.

[64] **Tekin, A. T., & Cebi, F.** (2020). Click and sales prediction for OTA's digital advertisements: Fuzzy clustering based approach. *Journal of Intelligent & Fuzzy Systems, 39* (5), 6619-6627.

[65] **Dorogoush, A. V., Ershov, V., & Gulin, A.** (2018). Catboost: Gradient boosting with categorical features. *ArXiv*, abs/1810.11363. p.1-7

[66] **Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y.** (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 3149–3157). Red Hook, NY, USA: Curran Associates Inc.

[67] **Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D., & Doyle, A. G.** (2018). Predicting reaction performance in C–N cross-coupling using machine learning. *Science 360* (6385), 186–190.

[68] **Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A.** (1984*). Classification and Regression Trees.* Boca Raton, FL: CRC press.

[69] **Bora, D. J., & Gupta A. K.** (2014). A comparative study between fuzzy clustering algorithm and hard clustering algorithm. *International Journal of Computer Trends and Technology (IJCTT), 10* (2).

[70] **Feng, D., Wang, W., Mangalathu, S., Hu, G., & Wu, T.** (2021). Implementing ensemble learning methods to predict the shear strength of RC deep beams with/without web reinforcements. *Engineering Structures, 235*. doi:10.1016/J.ENGSTRUCT.2021.111979

[71] **Freund, Y., & Shapire, R.** (1995). A Decision-Theoretic Generalization of Online Learning and Application to Boosting. In *Proceedings of the Second European Conference on Computational Learning Theory* (pp. 23-27).

[72] **Freund, Y., & Shapire, R.** (1999). A short ıntroduction to boosting. *Journal of Japanese Society for Artificial Intelligence, 14* (5), 771-780

[73] **Hu, G., & Kwok, K. C. S.** (2020). Predicting wind pressures around circular cylinders using machine learning techniques. *J Wind Eng Ind Aerodyn, 198,* 104099.

[74] **Lee, K., Jeong, H. O., Lee, S., & Jeong, W. K.** (2019). CPEM: Accurate cancer type classification based on somatic alterations using an ensemble of a random forest and a deep neural network. *Sci. Rep., 9* (1), 1–9.

[75] **MacQueen, J. B.** (1967). Some Methods for Classification and Analysis of Multi Variate Observations. In L. M. L. Cam, & J. Neyman, (Eds.), *Proc. of the fifth Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281-297). University of California Press.

[76] **Mirzakhanov, V. E.** (2020). Value of fuzzy logic for data mining and machine learning: A case study, *Expert Systems with Applications, 162.* doi: 10.1016/j.eswa.2020.113781

[77] **Sokolova, M., Japkowicz, N., & Szpakowicz, S.** (2006). Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In Sattar, A., Kang, B., (Eds.), *AI 2006: Advances in Artificial Intelligence. AI 2006. Lecture Notes in Computer Science* (Vol 4304). Berlin, Heidelberg: Springer. doi:10.1007/11941439_114

[78] **Triguero, I., del Río, S., López, V., Bacardit, J., Benítez, J. M., & Herrera, F.** (2015). ROSEFW-RF: The winner algorithm for the ECBDL'14 big data competition: an extremely imbalanced big data bioinformatics problem. *Knowl.-Based Syst., 87,* 69–7

[79] **Zhou, Z. H.** (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman and Hall

[80] **Zhou, Z. H.** (2015). Ensemble learning. *Encyclopedia Biometr, 411* (6).

[81] **Matloob, F., Ghazal, T. M., Taleb, N., Aftab, S., Ahmad, M., Khan, M. A., ... & Soomro, T. R.** (2021). Software defect prediction using ensemble learning: A systematic literature review. *IEEE Access*

[82] **Weeraddana, D., Khoa, N. L. D., & Mahdavi, N.** (2021). Machine learning based novel ensemble learning framework for electricity operational forecasting. *Electric Power Systems Research, 201,* 107477.

[83] **Lee, K., Laskin, M., Srinivas, A., & Abbeel, P.** (2021). Sunrise: A simple unified framework for ensemble learning in deep reinforcement learning. In *International Conference on Machine Learning* (pp. 6131-6141). PMLR.

[84] **Tyralis, H., Papacharalampous, G., & Langousis, A.** (2021). Super ensemble learning for daily streamflow forecasting: Large-scale demonstration and comparison with multiple machine learning algorithms. *Neural Computing and Applications*, *33*(8), 3053-3068.

[85] **Ihnaini, B., Khan, M. A., Khan, T. A., Abbas, S., Daoud, M. S., Ahmad, M., & Khan, M. A.** (2021). A smart healthcare recommendation system for multidisciplinary diabetes patients with data fusion based on deep ensemble learning. *Computational Intelligence and Neuroscience*, *2021*.

[86] **Azeez, N. A., Odufuwa, O. E., Misra, S., Oluranti, J., & Damaševičius, R.** (2021). Windows PE Malware Detection Using Ensemble Learning. In *Informatics* (Vol. 8, No. 1, p. 10). Multidisciplinary Digital Publishing Institute.

[87] **Zhang, Z., Mansouri Tehrani, A., Oliynyk, A. O., Day, B., & Brgoch, J.** (2021). Finding the Next Superhard Material through Ensemble Learning. *Advanced Materials*, *33*(5), 2005112.

[88] **Lu, Y., Zhang, Z., Shangguan, D., & Yang, J.** (2021). Novel Machine Learning Method Integrating Ensemble Learning and Deep Learning for Mapping Debris-Covered Glaciers. *Remote Sensing*, *13*(13), 2595.

[89] **Ghosh, A., Sumpter, B. G., Dyck, O., Kalinin, S. V., & Ziatdinov, M.** (2021). Ensemble learning and iterative training (ELIT) machine learning: applications towards uncertainty quantification and automated experiment in atom-resolved microscopy. *arXiv preprint arXiv:2101.08449*.

[90] **Chicco, D., & Jurman, G.** (2021). An ensemble learning approach for enhanced classification of patients with hepatitis and cirrhosis. *IEEE Access*, *9*, 24485-24498.

[91] **Yu, Z.**, **Lu, Y., Zhang, J., You, J., Wong, H., Wang, Y., & Han, G.** (2018). Progressive semisupervised learning of multiple classifiers. *IEEE Trans. Cybern. (48).* 689-702

[92] **Yu, Z.,Luo, P., & Liu, J.** (2018). Semi-supervised ensemble clustering based on selected constraint projection. *IEEE Trans. Knowl. Data Eng.(30).* 2394-2407.

[93] **Yu, Z., Zhang, Y., You, J., Chen, C.L.P., Wong, H., Han G., & Zhang, J.** (2019). Adaptive semi-supervised classifier ensemble for high dimensional data classification. *IEEE Trans. Cybern. (49).* 366-379.

[94] **Cui, S., Wang, Y., Yin, Y., Cheng, T.C.E., Wang, D., & Zhai, M**. (2021). A cluster-based intelligence ensemble learning method for classification problems. *Information Sciences (560).* 386-409

[95] **Hüllermeier, E**. (2015). Does machine learning need fuzzy logic?. *Fuzzy Sets Syst. (281)*. 292-299

[96] **Chan, P.P., Zheng, J., Liu, H., Tsang, E.C.C. & Yeung, D.S.** (2021). Robustness analysis of classical and fuzzy decision trees under adversarial evasion attack. *Applied Soft Computing*, *107*, p.107311.

[97] **Kumar, D.M., Satyanarayana, D. & Prasad, M.G.** (2021). MRI brain tumor detection using optimal possibilistic fuzzy C-means clustering algorithm and adaptive k-nearest neighbor classifier. *Journal of Ambient Intelligence and Humanized Computing*, *12*(2), pp.2867-2880.

[98] **Viji, C., Rajkumar, N., Suganthi, S.T., Venkatachalam, K. & Pandiyan, S.** (2021). An improved approach for automatic spine canal segmentation using probabilistic boosting tree (PBT) with fuzzy support vector machine. *Journal of Ambient Intelligence and Humanized Computing*, *12*(6), pp.6527-6536.

[99] **Bezdek J.**C. (1973). Fuzzy Mathematics in Pattern Classification.

CURRICULUM VITAE

Name Surname **: Ahmet Tezcan TEKİN**

EDUCATION **:**

- **B.Sc.** **:** 2013, Istanbul Technical University, Faculty of Computer and Informatics Engineering, Information Systems Engineering
- **B.Sc.** **:** 2013, Binghamton University, Faculty of Watson Engineering School, Information Systems Engineering
- **M.Sc.** **:** 2017, Istanbul Technical University, Management Faculty , Management Engineering Department

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2013-2017 Head of Data at Kontra Digital Services Inc.
- 2017-2020 CTO and R&D Manager at Cerebro Software Services Inc.
- 2020-…. Co-founder at Softtowel Games Inc.

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Tekin, A. T.**, & Cebi, F. (2020). Click and Sales Prediction for Digital Advertisements: Real World Application for OTAs. In: Kahraman, C., Cebi, S., Cevik Onar, S., Oztaysi, B., Tolga, A., Sari, I. (Eds.), *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making. INFUS 2019. Advances in Intelligent Systems and Computing* (Vol. 1029). Springer, Cham.
- **Tekin, A. T.**, & Çebi, F. (2020). Click and sales prediction for OTA's digital advertisements: Fuzzy clustering based approach. *Journal of Intelligent & Fuzzy Systems, 39* (5), 6619-6627
- **Tekin, A. T.**, Kaya, T., & Çebi, F. (2021). Click Prediction in Digital Advertisements: A Fuzzy Approach to Model Selection. In: Kahraman, C., Cevik, Onar, S., Oztaysi, B., Sari, I., Cebi, S., Tolga, A., (Eds.), *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions. INFUS 2020. Advances in Intelligent Systems and Computing* (Vol. 1197). Springer, Cham.

- **Tekin, A. T.**, Kaya, T., & Çebi, F. (2021). Customer lifetime value prediction for gaming industry: Fuzzy clustering based approach. *Journal of Intelligent & Fuzzy Systems, 42* (1), 87-96
- **Tekin, A. T.,** Çebi, F., & Kaya, T. (2022). Retention Prediction in the Gaming Industry: Fuzzy Machine Learning Approach. In Calisir, F., (Eds.), *Industrial Engineering in the Big Age of Business Intelligence*, Pre-press.
- **Tekin, A. T.,** Çebi, F., & Kaya, T. (2022). Stock Price Prediction: Fuzzy Clustering Based Approach. In: Wang, J., (Eds.), *Encyclopedia of Data Science and Machine Learning*, Pre-press. IGI-Global.

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- Öztayşi, B., **Tekin, A. T.**, Özdikicioğlu, C., & Tümkaya, K. C. (2017). Personalized Content Recommendation Engine for Web Publishing Services Using Textmining and Predictive Analytics. In R. Sahu, M. Dash, & A. Kumar, (Eds.), *Applying Predictive Analytics Within the Service Sector* (pp. 113-124). IGI Global.
- **Tekin, A. T.,** Ozkale, N. L., & Ayhan, G. (2018). The Importance of R & D Investments in Information and Communication Technologies and Tax Policies Applied through Information and Communication Technologies. In *Proceedings of the International R&D, Innovation and Technology Management Congress* (pp. 155-166).
- **Tekin, A. T.**, Ayhan, G., Ozkale, N. L. (2018). Turkish Automotive Industry and Preparation for Industry 4.0. In *International Engineering and Technology Management Summit* (pp. 1-8).
- Çakmak, T., **Tekin, A. T.,** Şenel, Ç., Çoban, T., & Şakar, C**.** (2019). Accurate Prediction of Advertisement Clicks Based on Impression and Click-Through Rate using Extreme Gradient Boosting. In *Proceedings of the 8th International Conference on Pattern Recognition Applications and Methods* (pp. 621-629).
- **Tekin, A. T.,** Ozkale, N. L., & Oztaysi, B. (2019). Big Data Concept in Small and Medium Enterprises: How Big Data Effects Productivity. In: Calisir, F., Cevikcan, E., Camgoz Akdag, H. (Eds.), *Industrial Engineering in the Big Data Era. Lecture Notes in Management and Industrial Engineering*. Springer, Cham.
- Tokuç, A. A., Uran, Z. E., & **Tekin, A. T.** (2019) Management of Big Data Projects: PMI Approach for Success. In Bolat, H. B., Temur, G. T. (eds) *Agile Approaches for Successfully Managing and Executing Projects in the Fourth Industrial Revolution* (pp. 279-293). IGI Global. doi:10.4018/978-1-5225-7865-9.ch015
- Şahinarslan, F.V, **Tekin, A. T.**, & Çebi, F. (2019). Machine Learning Algorithms to Forecast Population: Turkey Example. In *International Engineering and Technology Management Summit* (pp. 279-286).
- İdemen Tuğcu, B., **& Tekin, A. T.** (2019). Effects of Information Technologies Developments on Business Life Conditions. In *International Engineering and Technology Management Summit* (pp. 112-120).

- **Tekin, A. T.**, Özkale, L., & Gültekin-Karakaş, D. (2020). The Turkish Automotive Industry in the Era of Digital Technologies and Autonomous Cars. In Durakbasa, N., Gençyılmaz, M. (Eds.), *Proceedings of the International Symposium for Production Research 2019. Lecture Notes in Mechanical Engineering.* Springer, Cham.
- Özgün, K., Aklan, S. C., **Tekin, A. T.**, & Çebi, F. (2021). Malfunction Detection on Production Line Using Machine Learning: Case Study in Wood Industry. In Kahraman, C., Cevik Onar, S., Oztaysi, B., Sari, I., Cebi, S., Tolga, A. (Eds.), *Intelligent and Fuzzy Techniques: Smart and Innovative Solutions. INFUS 2020. Advances in Intelligent Systems and Computing* (Vol. 1197). Springer, Cham.
- Altınok, N., Farrokhizadeh, E., **Tekin, A. T.**, Khameneh, S. G., et al. (2022). Predicting Performance of Legal Debt Collection Agency. In Kahraman, C., Cebi, S., Cevik Onar, S., Oztaysi, B., Tolga, A. C., Sari, I. U. (Eds.), *Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation. INFUS 2021. Lecture Notes in Networks and Systems* (Vol. 308). Springer, Cham.
- Yağcıoğlu, E., **Tekin, A. T.**, & Çebi, F. (2022). Demand Forecasting of a Company that Produces by Mass Customization with Machine Learning. In: Kahraman, C., Cebi, S., Cevik Onar, S., Oztaysi, B., Tolga, A. C., Sari, I. U. (Eds.), *Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation. INFUS 2021. Lecture Notes in Networks and Systems* (Vol. 308). Springer, Cham.
- Şahinarslan, F. V, **Tekin, A. T.**, & Çebi, F. (2022). Machine Learning Approach to Forecast Population: Turkey Example. *International Journal of Data Science*. Pre-press.