|  |  |
| --- | --- |
| AKÜ FEMÜBİD **XX** (202X) xxxxxx (xx s) | AKU J. Sci. Eng. **XX** (202X) xxxxxx (xx pp) |

**DOI:**

**Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi**

**Afyon Kocatepe University Journal of Science and Engineering**

**Araştırma Makalesi / Research Article**

**Analysis of Factors Affecting Death Numbers: The Role of Other Factors Beyond Diseases**

**Ayşegül Sinem Karaman**

**1** *İstanbul Arel Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul.*

*e-posta: sinemkaramannn@gmail.com. ORCID ID:* [*http://orcid.org/0000-0001-1111-2222*](http://orcid.org/0000-0001-1111-2222)

Geliş Tarihi: ; Kabul Tarihi:

|  |  |
| --- | --- |
| ***Keywords***  ***Chronic Diseases; Death; CatBoost; RandomForest + GradientBoosting + K-Nearest Neighbors; K-Nearest Neighbors + XGBoost;*** | **Abstract** |
| There are some diseases in the world that have profound effects on global health and cause deaths. While these are chronic diseases such as heart diseases and cancer, the recent increase in infectious diseases such as Covid-19 has caused serious death consequences. Investigating these deaths and determining whether there are different reasons behind the deaths is an issue that needs to be emphasized for the human population. In this article, the relationship between the diseases mentioned and deaths in America is discussed, and it is aimed to evaluate the effects of the diseases on death rates by looking at the past origins of the people living there. Three different models were used for this study: (CatBoost), (RandomForest + GradientBoosting + K-Nearest Neighbors), and (K-Nearest Neighbors + XGBoost) models. |

**Ölüm Sayılarını Etkileyen Faktörlerin Analizi: Hastalıkların Ötesinde Diğer Etkenlerin Rolü**

|  |  |
| --- | --- |
| ***Anahtar kelimeler***  Kronik Hastalıklar; Ölüm; CatBoost; RandomForest + GradientBoosting + K-Nearest Neighbors; K-Nearest Neighbors + XGBoost; | **Öz** |
| Dünyada küresel sağlık üzerinde derin etkilere, ölümlere sebep olan bazı hastalıklar vardır. Bunlar Özellikle kalp hastalıkları ve kanser gibi kronik hastalıklar olurken son zamanlarda artan bulaşıcı hastalıklar örneğin covid -19 gibi ciddi anlamda ölüm sonuçlarına yol açmıştır. Bu ölümlerin araştırılması, ölümlerin arkasında yatan farklı sebeplerin olup olmadığının belirlenmesi insan popülasyonu için üzerinde durulması gerekilen bir konudur. Bu makale de belirtilen hastalıkların Amerika’da ölümler ile ilişki ele alınmış, orada yaşayan insanların geçmiş kökenlerine bakılarak ölüm oranları üzerinde ne kadar etkili olduğu sonuçları değerlendirilmek istenmiştir. Bu çalışma için üç tane farklı model kullanılmış olup bunlar (CatBoost), (RandomForest + GradientBoosting + K-Nearest Neighbors), ve (K-Nearest Neighbors + XGBoost) modelleridir. |

© Afyon Kocatepe Üniversitesi

# 

# 1. Introduction

Worldwide, health is one of the most critical areas required for a person to continue his life and perform his vital activities. Nowadays, various factors such as ethnicity, regional differences, gender and age group between countries can cause death as well as health problems.

(Bozkurt et al.2023). used in this article

The time spent by people in America in the dataset

It gives the number of people who died from diseases.If we talk about these diseases, the first one is sepsis. Sepsis disease; It is a disease known to cause organ dysfunction. It covers 19.7% of deaths worldwide. (Zhang et al.2022). Malignant neoplasms; It is a condition in which normally functioning cells are no longer able to function due to their proliferation more than normal. These cells are also called malignant cells (Solid et al. 2023).

Diabetes Mellitus; It is a disease that occurs when the pancreas cannot produce enough insulin and the body cannot use it (Chang et al., 2023). Alzheimer's disease; It is a leading cause of dementia worldwide. It is the extracellular accumulation of amyloid β peptides and the formation of intracellular neurofibrillary tangles (Uwişema et al. 2022).

Influenza and pneumonia; They are two different respiratory diseases. It has symptoms such as high fever, sore throat and cough (Owolabi et al. 2022).

Chronic lower respiratory tract diseases; They are serious respiratory diseases found in the elderly and adults. In the USA, this disease is known as a seasonal respiratory disease (Melgar et al. 2023).

Nephritis nephrotic syndrome And nephrosis; It is CNS caused by mutations in the NPHS1 (nephrin) gene. It usually causes severe proteinuria in the neonatal period. Heart diseases; It is a condition in which the heart cannot perform its functions. It generally affects the veins.

(TR et al.2022).

It is a disease that tags the function of blood vessels in the brain. (Nguyen et al. 2022).

Another disease we will examine is COVID-19; It is a type of infectious disease that affects the respiratory tract, especially in 2019. (Owolabi et al. 2022). Today, with the developing technology, machine learning is now used in almost every field. It has become popular in the field of health for many reasons such as examining diseases and diagnosing causes of death.

In this study, the effect of the number of deaths in America on diseases, as well as the effect of various factors on the number of deaths, were discussed using machine learning methods. If we talk about a few previous studies similar to this subject, a study was conducted on the diagnosis of diabetes using a machine learning model. K-EYK,C4.5 algorithm and RO models were used here. (Bölükbaşı 2023) Another study is the diagnosis of obesity disease with optimized supervised learning algorithms, where knn and random forest models were used. It was written to diagnose errors (Turan 2023). In general, studies were conducted on a single disease. In the study in this article, primarily data preprocessing stages were used. Later, various models were applied and these models were used as hybrids. These are: 'Sex' column is coded as 1 for 'M' value and 0 for 'F' value, 'AgeGroup' column is processed and coded categorically, 'Race/Ethnicity' column It is converted to numerical values with the Label Encoding method,

Numerical columns are normalized with Min-Max normalization. The number of disease deaths has been modeled according to several variables. Firstly, the ethnicity variable on the number of deaths from diseases was discussed. The effect of race/ethnicity on diseases was examined. The dependent variable was selected as ethnicity and the independent variables were set as all diseases.

Here it allows to create and train a classification model using the catboost library. The accuracy result was found to be 0.82 here.

The next issue addressed is the effect of gender variable on the number of deaths from diseases. What is the effect of gender on causes of death here? We wanted to find an answer to this question. Here, gender is given as the dependent variable and independent variables are set as diseases that cause death. Here, a hybrid model has been created by combining Random Forest, Gradient Boosting and K-Nearest Neighbors. Here, the effect of gender on the number of deaths was examined and the accuracy result was found to be 0.87. Then, according to the results obtained, the disease that caused the most deaths in men and women was discussed and Diabetes mellitus was found to be the disease that caused the most deaths for both gender variables. The next issue discussed was the effect of age groups on the number of deaths from diseases. Age groups were considered categorically on the dataset used, but were converted to numerical values for the application. Here, KNN and XGBoost models were used together and the Voting Classifier model was created. Here, an accuracy of 0.84 was taken. Then, a bar chart was created to see the effect of death numbers on age groups. Then, time series graphs, confisuon matrix and correlation analyzes were made to support the results.

|  |  |
| --- | --- |
| Dataset Variables | Variable Description |
| AnalysisDate | Date of analysis |
| Date Of Death Year | year of death |
| Date Of Death Month | death moon |
| Start Date | Research start date |
| End Date | Research end date |
| Jurisdiction of Occurrence | Where is it made? |
| Sex | Gender |
| Race/Ethnicity | Ethnicity/Race |
| AgeGroup | Age group |
| AllCause | Number of deaths from all causes |
| NaturalCause | Number of people who died naturally |
| Septicemia (A40-A41) | Number of deaths from septicemia |
| Malignant neoplasms (C00-C97) | Number of deaths from malignant neoplasms |
| Diabetes mellitus (E10-E14) | Diabetes Mellitus death toll |
| Alzheimer disease (G30) | Alzheimer's death toll |
| Influenza and pneumonia (J09-J18 | lower respiratory tract deaths |
| Chronic lower respiratory diseases (J40-J47 Other diseases of respiratory system (J00-J06,J30-J39,J67,J70-J98) | Number of deaths from chronic respiratory disease |
| Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27) | Number of people dying from nephritis, nephrotic syndrome and nephrosis |
| Symptoms, signs and abnormal clinical and laboratory findings not elsewhere classified (R00-R99) | Number of deaths from disease with symptoms, signs, and abnormal clinical and laboratory findings |
| Diseases of heart (I00-I09,I11,I13,I20-I51) | Number of people dying from heart disease |
| Cerebrovascular diseases (I60-I69) | Number of deaths from cerebrovascular disease |
| COVID-19 (U071, Multiple Cause of Death) | Number of people who died from COVID-19 disease |
| COVID-19 (U071, Underlying Cause of Death) | Number of people who died from COVID-19 disease |

**Table 1. Data Set Table**

**2. Data Set Used in the Study**

The data set used for the study was taken from the data.gov website and gives us the relationship between diseases and the number of deaths. There are 23 variables used in the data set; Date of Analysis, Date of Death Year, Date of Death Month, Start Date, End Date, Site of Occurrence, Gender, Race/Ethnicity, Age Group, All Causes, Natural Cause, Septicemia (A40-A41), Malignant neoplasms (C00-C97) ), Diabetes Mellitus (E10-E14),Alzheimer's disease (G30),Flu and pneumonia (J09-J18),Chronic lower respiratory tract diseases (J40-J47),,"Other diseases of the respiratory system (J00,J06,J30) -J39 ,J67,J70))-J98)" Nephritis, nephrotic syndrome and nephrosis (N00-N07,N17-N19,N25-N27)""Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified (R00-R99) ) " Heart diseases (I00-I09,I11,I13,I20-I51)",Cerebrovascular diseases (I60-I69),"COVID-19 (U071, Multiple Causes of Death)","COVID-19 (U071, Main Cause) Death ) )" variables. Place of Formation, Gender, Race/Ethnicity, Age Group variables are categorical variables, and the remaining variables contain numerical data. The data set contains a total of 3961 data. Studies were carried out over a certain period of time to collect this data. The results of each disease-related death were collected at certain periods in 2019, and the analyzes were carried out in 2021. These data were collected by the US Department of Health and Human Services Data and are a highly accurate data set. In order to provide easier access to these data, they are presented as a table in the appendix of the article.

**3) Pre-Application Data Processing**

Before starting model training, the data was put into pre-processing processes.

***3.1 Editing Date Information***

We process date data using AnalysisDate, Date Of Death Year, Date Of Death Month, Start Date, and End Date columns. Using the Start Date and End Date columns, the number of days between the start and end dates of the analysis is found. We create a column called DeathDuration and write the number of days in this column. The purpose of doing this is to investigate whether the time spent in the correlation analysis has an effect on the number of deaths and to scale which diseases it is effective in.

***3.2 Converting Categorical Values to Numerical***

In the project, the categorical 'Sex', 'AgeGroup' and 'Race/Ethnicity' columns in the dataset are converted to numerical values for processing. In the Sex column, numerical values are given to women as 1 and to men as 0, and in the AgeGroup column,

**Figure 1.** Correlation Analysis

unique numbers are assigned. As for the Race/Ethnicity column, label encoding is applied and a new column is created and the old column is deleted from the data set***.***

***3.3 Correlation Analysis***

It was used to examine the correlation between the number of days in the day column in which we found the death events and the number of deaths caused by various diseases.

***3.4 Control of Missing and Unnecessary Data in the Data Set***

Missing data in the data set we use is checked and if there is missing data, it is filled with the average value. Then, unnecessary data; Columns 'Jurisdiction of Occurrence', 'AnalysisDate' are removed.

***3.5 Normalization Procedures***

Here, min-max normalization process is applied to the numerical values in the data set, so that the numerical values ​​must fall between 0 and 1. The developed models should give more accurate results because the number of deaths consists of very high numbers.

**4. Machine Learning Algorithms**

In this section, the machine learning models used in the study are mentioned.metin, ekran görüntüsü, paralel içeren bir resim

Açıklama otomatik olarak oluşturuldu

***4.1 CatBoost Classifier***

CatBoost is a system that can handle categorical variables efficiently and is known for its fast training times.Is the model. With the ability to effectively handle categorical variables and parallel processing capabilities improves performance on large data sets. Since it automatically determines the learning rate, it greatly reduces overfitting (Rathod et al. 2022). CatBoost can be used for both regression and classification problems. If we explain it numerically;

Model: Here, M represents the total number of trees and (x) represents the prediction of each tree.

(1)

Tree Estimation: w refers to the weight vector, and q(x) refers to the tree nodes selected from the input features x.

 (2)

Loss Function: It solves the difference between the actual value and the estimated value.

(3)

Tree Complexity: Cost is checked here**.**

(4)

In this study, 'Race/Ethnicity' was used as the target and all disease types were used as features. Then, the data set was divided into training and test sets using the train\_test\_split function. A model was created using the CatBoostClassifier class. The basic parameters of the model were defined as parameters such as the number of trees, tree depth, and learning rate.

The model we created with the fit method is trained. The eval\_set parameter was used to evaluate the performance of the model during the training process. The accuracy rate (acc value) of the model was measured as 0.82.

***4.2 Ensemble Model***

Ensemble methods are learning algorithms that classify new data by creating a set of classifiers and then taking the results of their predictions. These models can give higher prediction performance results because they increase their complexity and diversity. (Rezaei 2023)

These models are examined in two basic categories, the first of which is;

Bagging (Bootstrap Aggregating): Trained on different subsamples and then aggregated.

Boosting: In each basic example, it tries to move forward by correcting previous errors.

Mathematically speaking, the model prediction is expressed as a combination of the prediction of the created models.

(5)

In this study, as an ensemble model; Random Forest, K-Nearest Neighbor, Gradient Boosting models were applied. If we talk about these models;

***4.2.1 Random Forest***

It is a classification and regression algorithm used in ensemble modeling technique. Random Forest allows combining many decision trees to create a powerful model. (Ozkan 2023).

In the ensemble model in this article, first

n\_estimators variable is the number of trees to be created,

The seed value is used for randomness control with the random\_state variable, and the class\_weight='balanced' value is used to handle the imbalance between classes for datasets with unbalanced classes.

**ekran görüntüsü, yazı tipi, grafik, metin içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**Figure 2.**

***4.2.2 K-Nearest Neighbor***

It is an easy and effective model. Its basic principle is the method in which the value of a data point is determined according to the values of its neighbors. The nearest neighbor is taken as the basis. (Filiz 2023).

If we put it mathematically;

(6)

It can be expressed as:

In this project, the KNeighborsClassifier classifier model was used with the knn\_model parameter.

Later, the knn model was used in VotingClassifier along with other models.

metin, diyagram, yazı tipi, daire içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Figure 3.**

***4.2.3 Gradient Boosting***

Each decision tree estimates the error of the previous decision tree, increases and improves the error as a gradient (Malik et al. 2022).

(7)

In this formula;

F(x) = represents the total estimate.

F0(x) = prediction of the initial model.

h1(x),h2(x),…,hr(x)= are the predictions of the models added respectively.

metin, yazı tipi, çizgi, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Figure 4.** Gradient boosting algorithm

In this study, the GradientBoostingClassifier classifier model is defined with the gb\_model parameter. And then, these three models were trained by using them in VotingClassifier together with other models (randomforest, knn) and predictions were made on the trained data set and their performance was evaluated in comparison with the real value. In this project, the sex column, which is a categorical value, was chosen as the target variable, and the variables giving the number of all diseases that cause death were used as independent variables. It was tried to obtain the results of whether gender has an effect on the number of deaths, and if so, what is the effect of this rate on women and men, and what is the disease that causes the most deaths in men and women. Some values measured in these models:

• Most common cause of death for women: NaturalCause

• The most common cause of death for men: NaturalCause It was measured as .

***4.3 VotingClassifier*** ***Model***

This model is the other and last model used in our study. Here too, two models were used in hybrid form.This strategy combines the predictions of classification models with a specific weighting. It uses the probability values produced by the models it combines and obtains a total probability score by processing these probability values. Then, a prediction is made based on this probability score. (Rai et al. 2022) In this study, knn and xgboost models were combined into a votingclassifier model to obtain a result.

***4.3.1 K-Nearest Neighbor***

A classifier model was created taking into account the 5 nearest neighbors at the location specified with n\_neighbors=5. It was later added to the VootingClassifier model to be evaluated as a community with the XGBoot model.

***4.3.2 XGBoost Model***

XGBoost, an ensemble tree algorithm developed by Chen and Guestrin, is an efficient model used to produce a hybrid model whose prediction performance is better than the techniques used alone. (Tokmak 2023) Mathematically, its basic formula is shown as follows.

(8)

Here two variables represent the estimated model.

metin, ekran görüntüsü, logo, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Figure 5.** XGBOOST model

In this article, it was combined with knn and added into a VotingClassifier Model and the training phase was started. Predictions were made based on the trained values. The target variable was determined as the AgeGroup variable, and the independent variables were selected as all diseases given by the number of deaths that cause the disease. Thus, it was aimed to make a classification estimate on the number of deaths from diseases in the age group. The accuracy rate was measured as 0.84.

**5.** **Results Obtained from Applied Models**

As a result of the research and applications, the precision, recall, F-1 core, support and accuracy rates of the applied models are as follows.

***5.1 Catboost Classifier Results***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | precision | recall | F1-score | support |
| 0 | 0.87 | 0.87 | 0.87 | 129 |
| 1 | 0.70 | 0.80 | 0.75 | 128 |
| 2 | 0.80 | 0.80 | 0.80 | 143 |
| 3 | 0.92 | 0.86 | 0.89 | 125 |
| 4 | 096 | 0.95 | 0.96 | 124 |
| 5 | 0.72 | 0.66 | 0.69 | 143 |

***Table 2.*** Catboost classifier results are as seen in the table.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Accuracy |  |  | 0.82 | 792 |
| Macro avg | 0.83 | 0.83 | 0.83 | 792 |
| Weighted avg | 0.82 | 0.82 | 0.82 | 792 |

***Table 3.*** Catboost classifier results are as seen in the table.

***5.2 Ensemble Model Results***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | support |
| 0.0 | 0.87 | 0.98 | 0.92 | 628 |
| 1.0 | 0.87 | 0.41 | 0.56 | 164 |
| Accuracy |  |  | 0.87 | 792 |
| Macro avg | 0.87 | 0.70 | 0.74 | 792 |
| Weighted avg | 0.87 | 0.87 | 0.85 | 792 |

***Table 4.*** Ensemble Model results are as seen in the table.

|  |  |
| --- | --- |
| Ensemble Model Acc | 0.8661616161616161 |
| Kadınları en çok öldüren Hastalık | NaturalCase |
| Erkekleri en çok öldüren hastalık | NaturalCase |

***Table 5.*** Ensemble Model results are as seen in the table.

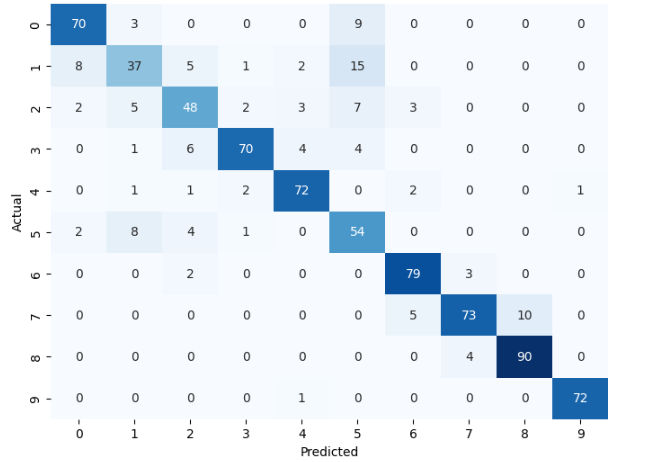
## *5.3. VotingClassifier* *Model Results*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | Recall | F1-score | support |
| 0 | 0.85 | 0.85 | 0.85 | 82 |
| 1 | 0.67 | 0.54 | 0.60 | 68 |
| 2 | 0.73 | 0.69 | 0.71 | 70 |
| 3 | 0.92 | 0.82 | 0.87 | 85 |
| 4 | 0.88 | 0.91 | 0.89 | 79 |
| 5 | 0.61 | 0.78 | 0.68 | 69 |
| 6 | 0.89 | 0.97 | 0.91 | 84 |
| 7 | 0.91 | 0.83 | 0.87 | 88 |
| 8 | 0.90 | 0.96 | 0.93 | 94 |
| 9 | 0.99 | 0.99 | 0.99 | 73 |
| Accuracy |  |  | 0.84 | 792 |
| Macro avg | 0.83 | 0.83 | 0.83 | 792 |
| Weighted avg | 0.84 | 0.84 | 0.84 | 792 |

***Table 5***. VotingClassifier Model results are as shown in the table.

|  |  |
| --- | --- |
| Voting Model Accuracy | 0.8396464646464646464 |

***Tablee 6.*** VotingClassifier Model results are as shown in the table.



**Figure 9.** Configuration matrix result of the votingclassifier model.

**ekran görüntüsü, metin, dikdörtgen, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**Figure 10.** A graphical representation of the agegroup variable is given. Age groups are on the x-axis and estimated ages are on the y-axis..

**6.Model Comparisons**

In the previous section, a graphic comparison of the accuracy rates of the three different results we obtained using different classification models was made. The classification models used are specified, each model is created as a series, the values are read respectively and shown in the chart below.

ekran görüntüsü, metin, dikdörtgen, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Figure 10.** Model accuracy comparison

According to the results obtained in this graph, the ensemble model gave more accurate results.

**7.Evaluation Criteria**

These are the evaluation criteria by which we get the results after training our model and making predictions and learn how accurate the prediction is.

***7.1 Precision***

Precision is a value that expresses the ratio of values predicted as positive by a classification model to values that are actually positive. (Oikonomou et al. 2023)

(9)

True Positives (TP): These are the cases where the model correctly predicts positive.

False Positives (FP): These are cases where the model incorrectly predicts a positive.

The precision value is a number between 0 and 1. The closer it is to 1, the more positive the model's predictions are. (Miao et al. 2023)

***7.2 Recall***

Recall is a performance metric that shows how many true positive values a classification model correctly detects.

Here too, the values take values between 0 and 1. The closer they are to 1, the more accurate the predictions are. ( . Banerjee et al.2023)

***7.3 F-1 Score***

F1-score is the harmonic mean of precision and recall. By combining these two criteria, F1-score allows to summarize the performance of a model as a single value.

metin, yazı tipi, beyaz, tasarım içeren bir resim

Açıklama otomatik olarak oluşturuldu (10)

***7.4 Support***

It represents the number of times any given class appears in the actual data set. Classification reports usually give the number of samples reported separately for each class. This support value is also taken into consideration when calculating metrics such as precision, recall, and F1-score. (Kufel et al. 2023)

***7.5 Confision Matrix***

True Positive (TP): It is a true positive model. The situation that the model correctly predicts as positive

True Negative (TN): It is called the true negative model. The situation that the model correctly predicts as negative

False Positive (FP): It is a false positive model. It is a situation that the model predicts as positive but is actually negative.

False Negative (FN): It is a false negative model. It gives the situation that the model predicts as negative but is actually positive. (Febrian et al.2023)

metin, ekran görüntüsü, yazı tipi, sayı, numara içeren bir resim

Açıklama otomatik olarak oluşturuldu

***Figure 12.*** Visual representation of confusion matrix value

***7.6 Time Series Analysis***

With our latest results, we analyze the data by year and create a time graph to find the total number of deaths. Here, y variables are the grouped death numbers and x variable is the year of death indicator.

**çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram, eğim, bayır içeren bir resim

Açıklama otomatik olarak oluşturuldu**

**Şekil 13** Time series analysis grouping numbers by year of death

**8. Discussion and Conclusion**

In this article, the data of people who died due to various diseases in America was processed. A classification process was applied on numerical data with certain categorical values. Models were trained and then predictions were made. With the predictions made, certain categorical values (AgeGroup, Sex, Race/Ethnicity columns were selected as target variables, and all columns giving the death numbers of other diseases mentioned in the data set were selected as independent variables, and how effective the target variables were on the independent variables was discussed. Each education and Different models were used for prediction and care was taken to use hybrid models. According to the results, the accuracy rates were measured between 0.80 and 0.90. Certain studies on this subject have previously been conducted on this subject, specializing only on a single disease rather than on all diseases; diabetes and obesity are just a few of them. However, this study focused on the variables on which the number of deaths from all diseases depend. This study can be used by a state's ministries, states, etc. that deal with deaths, and can provide detailed information about the people who died in that region. In fact, the health of people who died due to factors such as infectious diseases, which are still effective today. Their demographics can provide detailed information about the reasons behind their deaths. By conducting more in-depth analyzes in the future, it will open up advanced engineering opportunities for hospitals, relevant government ministries, institutions and organizations dealing with infectious diseases, providing an opportunity for developments in diseases that cause death, and perhaps more.

**9. Resources**

BOZKURT, İ. (2023). GEÇMİŞTEN GÜNÜMÜZE ÜLKEMİZİN PANDEMİLERE YÖNELİK SAĞLIK POLİTİKALARINA GENEL BİR BAKIŞ. Atlas Sosyal Bilimler Dergisi, 1(13), 51-66.

Zhang, W., Fang, X., Gao, C., Song, C., He, Y., Zhou, T., Yang, X., Shang, Y., Xu, J. (2023). MDSCs in sepsis-induced immunosuppression and its potential therapeutic targets. Journal Name, 5 June 2022,

Soldi, LR, Silva, VLC, Rabelo, DH ve diğerleri. Malign neoplazmların mikroçevresinde doğal öldürücü hücrelerin monoklonal antikorlarla yeniden aktivasyonu. J Cancer Res Clin Oncol 149 , 6661–6673 (2023).

Chang, V., Bailey, J., Xu, QA ve diğerleri. Makine öğrenimi (ML) algoritmalarına dayalı Pima Kızılderilileri diyabet sınıflandırması. Nöral Hesaplama ve Uygulama 35 , 16157–16173 (2023).

Tokito T, Kido T, Muramatsu K, Tokutsu K, Okuno D, Yura H, Takemoto S, Ishimoto H, Takazono T, Sakamoto N, ve diğerleri. İnfluenza Virüsü Pnömonisi İçin Hastaneye Yatıştan Sonraki 7 Gün İçinde İntravenöz Azitromisinin Uygulanmasının Etkisi: Ülke Çapında İdari Veri Tabanı Kullanan Bir Eğilim Skoru Analizi. Virüsler . 2023; 15(5):1142.

Uwişema, Ö. , Mahmud, Bir güneş , J. , Correia, IFS , Bejjani, N. , Alwan, M. , Nicholas, A. , Ölüyemi, A. ve Dost, B. ( 2022 ). Alzheimer hastalığı bulaşıcı bir nörolojik hastalık mıdır? Literatürün gözden geçirilmesi . Beyin ve Davranış , 12 , e2728.

Owolabi, M., Ali, R., Dacosta, J., Muhanna, A., & Slim, J. (2022). When Influenza, Bacterial Pneumonia, and COVID-19 Co-exist. Cureus, 14(12).

Melgar M, Britton A, Roper LE, et al. Yaşlı Yetişkinlerde Solunum Sinsityal Virüs Aşılarının Kullanımı: Bağışıklama Uygulamaları Danışma Komitesinin Önerileri — Amerika Birleşik Devletleri, 2023. MMWR Morb Mortal Wkly Rep 2023;72:793–801.

Jalanko, H., Jahnukainen, T., Ng, KH (2022). Konjenital Nefrotik Sendrom. İçinde: Emma, F., Goldstein, SL, Bagga, A., Bates, CM, Shroff, R. (eds) Pediatrik Nefroloji. Springer, Cham.

TR, R. ., Lilhore, UK, M, P. ., Simaiya, S. ., Kaur, . A. . ve Hamdi, M. . (2022). MAKİNE ÖĞRENME YAKLAŞIMLARIYLA KALP HASTALIKLARININ ÖNGÖRÜLEBİLİR ANALİZİ. Malezya Bilgisayar Bilimleri Dergisi , 132–148.

Nguyen, AT, Kouri, N., Labuzan, SA ve diğerleri. MRI'daki difüzyon değişiklikleriyle ilişkili serebrovasküler hastalığın nöropatolojik ölçekleri. Acta Neuropathol 144 , 1117–1125 (2022).

KABA, G. ve BAĞDATLI KALKAN, S. (2022). KARDİYOVASKÜLER HASTALIK TAHMİNİNDE MAKİNE ÖĞRENMESİ SINIFLANDIRMA ALGORİTMALARININ KARŞILAŞTIRILMASI. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, 21(42), 183-193.

Bölükbaşı, İ. B. (2023). *Dengesiz Bir Diyabet Veri Setinde Makine Öğrenmesi Yöntemlerini Kullanarak Diyabet Hastalığının Teşhisi* (Doctoral dissertation, Bursa Uludag University (Turkey)).

Y. Rathod ve diğerleri , "CatBoost Algoritmasını Kullanarak Polikistik Over Sendromunun Tahmin Edici Analizi", 2022 IEEE Bölge 10 Sempozyumu (TENSYMP) , Mumbai, Hindistan, 2022, s. 1-6, doi: 10.1109/TENSYMP54529.2022.9864439.

Wang, Y., Wang, R., Wang, J. ve diğerleri. Bayesian Optimizasyon Catboost Algoritmasına Dayalı Dalga Hızı ve Operasyonel Parametreleri Bütünleştiren Bir Kaya Kütlesi Dayanımı Tahmin Yöntemi. KSCE J Civ Eng 27 , 3148–3162 (2023).

Baradaran Rezaei, H., Amjadian, A., Sebt, MV *ve diğerleri.*Mide kanserinin prognozunu tahmin etmek için makine öğreniminin birleştirilmiş bir yöntemi. *Ann Oper Res*328 , 151–192 (2023).

ÖZKAN, B., PARİM, C. ve ÇENE, E. (2023). Ülkelerin Gelişmişlik Düzeylerinin Karar Ağacı ve Rastgele Orman Yöntemleriyle Tahmin Edilmesi. EKOIST Ekonometri ve İstatistik Dergisi ( 38 ), 87-104.

FİLİZ, E. (2022). Türkiye Covid-19 günlük hasta sayısındaki değişimin sınıflandırılmasına yönelik tahmininin destek vektör makineleri ve k-en yakın komşu algoritmaları ile gerçekleştirilmesi. Gümüşhane Üniversitesi Fen Bilimleri Dergisi, 12(1), 370-379.

Malik, A., Saggi, M. K., Rehman, S., Sajjad, H., Inyurt, S., Bhatia, A. S., ... & Yaseen, Z. M. (2022). Deep learning versus gradient boosting machine for pan evaporation prediction. Engineering Applications of Computational Fluid Mechanics, 16(1), 570-587.

Rai, N., Kaushik, N., Kumar, D., Raj, C. ve Ali, A. (2022). Yumuşak oylama sınıflandırıcısını kullanarak COVID-19 hastalarının ölüm tahmini. Uluslararası Mühendislikte Bilişsel Hesaplama Dergisi , 3 , 172-179.

TOKMAK, M. (2023). Öğrencilerin Siber Güvenlik Farkındalık Düzeylerinin Makine Öğrenmesi Yöntemleri ile Belirlenmesi. Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 28(2), 451-466.

*Oikonomou, EK, Khera, R. Hassas diyabet bakımı ve kardiyovasküler risk tahmininde makine öğrenimi. Cardiovasc Diabetol 22 , 259 (2023).*

*Miao, J., Zhu, W. Hassas hatırlama eğrisi (PRC) sınıflandırma ağaçları. Evrim. Intel. 15 , 1545–1569 (2022).*

*Banerjee, A., Mutlu, O. C., Kline, A., Surabhi, S., Washington, P., & Wall, D. P. (2023). Training and profiling a pediatric facial expression classifier for children on mobile devices: machine learning study. JMIR formative research, 7, e39917.*

*Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., ... & Gruszczyńska, K. (2023). What is machine learning, artificial neural networks and deep learning?—Examples of practical applications in medicine. Diagnostics, 13(15), 2582.*

*Febrian, M. E., Ferdinan, F. X., Sendani, G. P., Suryanigrum, K. M., & Yunanda, R. (2023). Diabetes prediction using supervised machine learning. Procedia Computer Science, 216, 21-30.*

TURAN, T. Optimize Edilmiş Denetimli Öğrenme Algoritmaları ile Obezite Analizi ve Tahmini. *Mehmet Akif Ersoy Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *14*(2), 301-312.