**VIOLENCE IN WOMEN AND TURKEY**

Table of Contents

[1 - Choosing and Proposing a Problem 2](#_Toc122992350)

[1.1 - General Information Company 2](#_Toc122992351)

[1.2 - Description of The Problem and Current Situation 2](#_Toc122992352)

[2 - Requirements Analysis 2](#_Toc122992353)

[2.1 - What the users expect from the developed system 2](#_Toc122992354)

[2.2 -Which type of questions do they need to be answered by means of the system 2](#_Toc122992355)

[2.3 - Conceptual Modeling of your Data Warehouse 3](#_Toc122992356)

[3-ETL Operations and KDD Process 3](#_Toc122992357)

[3.1- Descriptions of the ETL operations 4](#_Toc122992358)

[3.2 -Provide brief information about your Input Data 4](#_Toc122992359)

[3.3 -Use a Data Pre-Processing Method 5](#_Toc122992360)

[4- Mining and Prediction: 8](#_Toc122992361)

[4.1 -Data Mining Methods : 8](#_Toc122992362)

[4.2-Prediction and Methods: 11](#_Toc122992363)

# 1 - Choosing and Proposing a Problem

## - General Information Company

We found the dataset from Kaggle link which is provided by our instructor. The title of the Kaggle project is “Violence Against Women and Girls” (Larxel,2020). Kaggle writer takes the data from The Demographic and Health Surveys (DHS) and combine them country by country with different sociodemographic questions and answers. If there are more data to be needed in the project, we will provide with the program called DHS Program STATcompiler carried out by DHS.

## - Description of The Problem and Current Situation

Violence against women and girls is maybe the biggest issue in the world. Some countries are trying to reduce the violence rate; some have already fixed their violence rate at a low level. We are wondering how much our country’s violence rate and how much it differs from other countries. As we know, European countries and USA focus more on this issue and they are way ahead of us. We will make a comparison between Turkey, African countries, Asian countries and South American countries. Furthermore, we can make some conclusions such as “‘X’ country applies the ‘Y’ and their violence rate reasonably low level when we compare with our country. Maybe we can try to apply a similar approach.”

# 2 - Requirements Analysis

## 2.1 - What the users expect from the developed system

The effect of the factors in the processed data sets on violence against women in different countries and the place of Turkey in this situation. We can easily compare these countries and make a consequence of their values and according to their values, we can search for their approaches to this issue.

## 2.2 -Which type of questions do they need to be answered by means of the system

Respondents were asked if they agreed with the following statements:

-A husband is justified in hitting or beating his wife if she burns the food

-A husband is justified in hitting or beating his wife if she argues with him

-A husband is justified in hitting or beating his wife if she goes out without telling him

-A husband is justified in hitting or beating his wife if she neglects the children

-A husband is justified in hitting or beating his wife if she refuses to have sex with him

-A husband is justified in hitting or beating his wife for at least one specific reason

From these types of questions in our data set, we can examine the rate of violence against women and girls. Thanks to this system, people will be able to compare the rates of violence against women in their own countries with the rates in other countries, thereby creating awareness of where they are, and how improvements can be made, and prejudices against other countries can be broken.

## 2.3 - Conceptual Modeling of your Data Warehouse

We thought that the star schema would be more logical due to the small number of columns in the data set. Since our columns do not need a detailed explanation, they will often have their primary keys from our dimension table. The fact table will contain the collection of these primary keys and the measures of our value column.

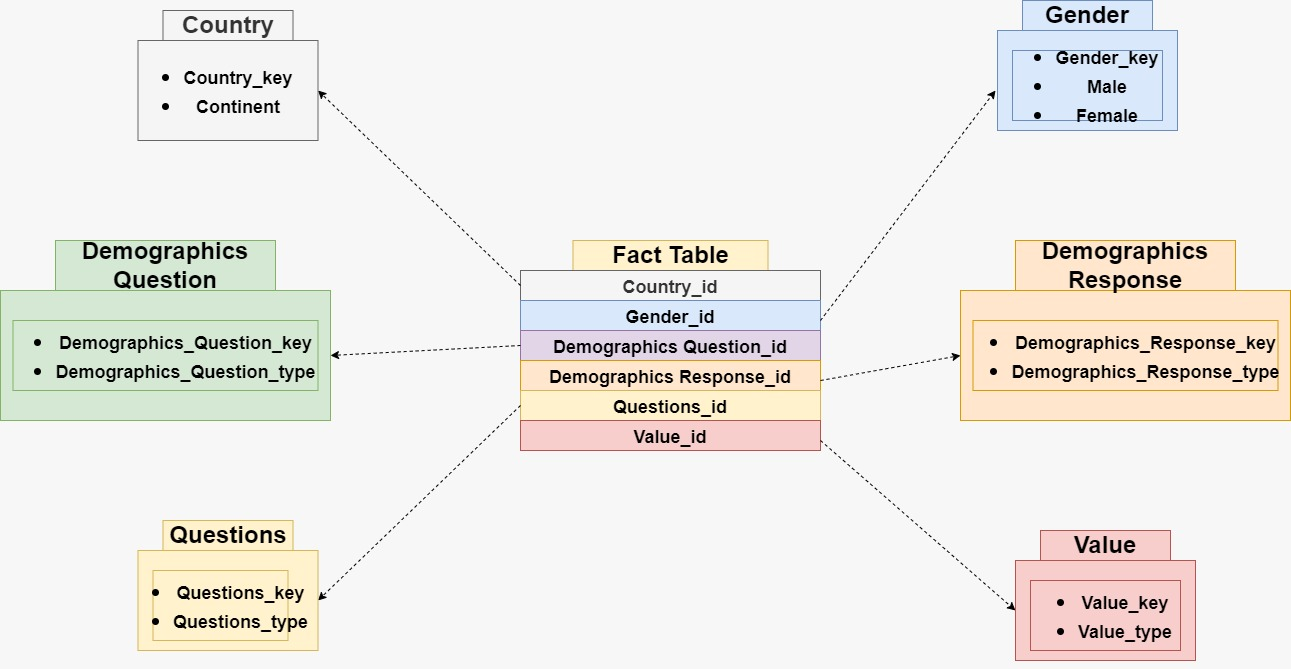


Figure 1 Conceptual Modeling

# 3-ETL Operations and KDD Process

## 3.1- Descriptions of the ETL operations

We used the dataset violence against women and girls from Kaggle that we mentioned in the first step, and since this dataset is very comprehensive and there are many observations from different countries, we did not need an extra dataset for the **extraction** stage.

Our data was generally compatible within itself, so we removed survey year and record ID dimension tables from the dataset in order to reduce the curse of dimensionality during the **transformation** phase, as they did not have a role in the project we were aiming for.

Since we are going through a single dataset, we use the Pandas library of the Python library instead of SQL to create our csv dataset; We have performed our cleaning, transformation and **loading** operations here.

## 3.2 -Provide brief information about your Input Data

There are 12,600 observations in total in our data set.

There are 5 different categories of questions from 3 unique continents and 70 different countries, and each question has 3 different answers. There are answers to 6 different questions.

Columns: Country, Gender, Demographic Questions, Demographic Response, Questions, Value

Country samples: 'Afghanistan', 'Albania', 'Angola', 'Armenia', 'Azerbaijan’, ‘Bangladesh', 'Benin', 'Bolivia', 'Burkina Faso', 'Burundi', 'Cambodia', 'Cameroon', 'Chad', 'Colombia', 'Comoros', 'Congo'…., ‘Turkey.

Gender samples: ‘Male’, ‘Female’.

Demographic Questions: 'Marital status', 'Education', 'Employment', 'Age', 'Residence'

Demographic Response sample: Never married', 'Higher', 'Secondary', 'Primary’, ‘Widowed, divorced, separated', 'Employed for kind', '15-24', 'Unemployed', 'Rural', '25-34', 'Married or living together’, ‘Urban', '35-49', 'No education', 'Employed for cash'

Questions:

-A husband is justified in hitting or beating his wife if she burns the food

-A husband is justified in hitting or beating his wife if she argues with him

-A husband is justified in hitting or beating his wife if she goes out without telling him

-A husband is justified in hitting or beating his wife if she neglects the children

-A husband is justified in hitting or beating his wife if she refuses to have sex with him

-A husband is justified in hitting or beating his wife for at least one specific reason

## 3.3 -Use a Data Pre-Processing Method

There were survey year and record id dimensions in the data we found and wanted to use. In order to keep the most recent data we can find and not to depend on the record IDs (we have already removed the dimensions that we do not need while data mining).

tablo içeren bir resim

Açıklama otomatik olarak oluşturuldu

tablo içeren bir resim

Açıklama otomatik olarak oluşturulduFigure 2 Dataset without changing

Figure 3 Dataset with Changing

If we look at the tables with the data, the data set seen in Figure [2] is the raw data we got from Kaggle. Since this raw data is not both record id and survey year dimensions related to the results we want, we deleted these two columns. We have converted it to the one seen in Figure [3].

At the same time, we reduced duplicate values and cleaned up unnecessary observations. There were too many missing values in our data, so in order to fix these missing values, we filled attribute mean for all samples based on other observations instead of missing values, so we used data cleaning methods in cleaning the data containing missing values.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Figure 4 Data Cleaning Code

As we see in Figure [4] We found the most similar person or people according to the values ​​in the columns and assigned the values ​​according to their averages. If the people we found also had null values, we assigned our values, downgrading the dimensions one by one in order of importance.

# 4- Mining and Prediction:

## 4.1 -Data Mining Methods :

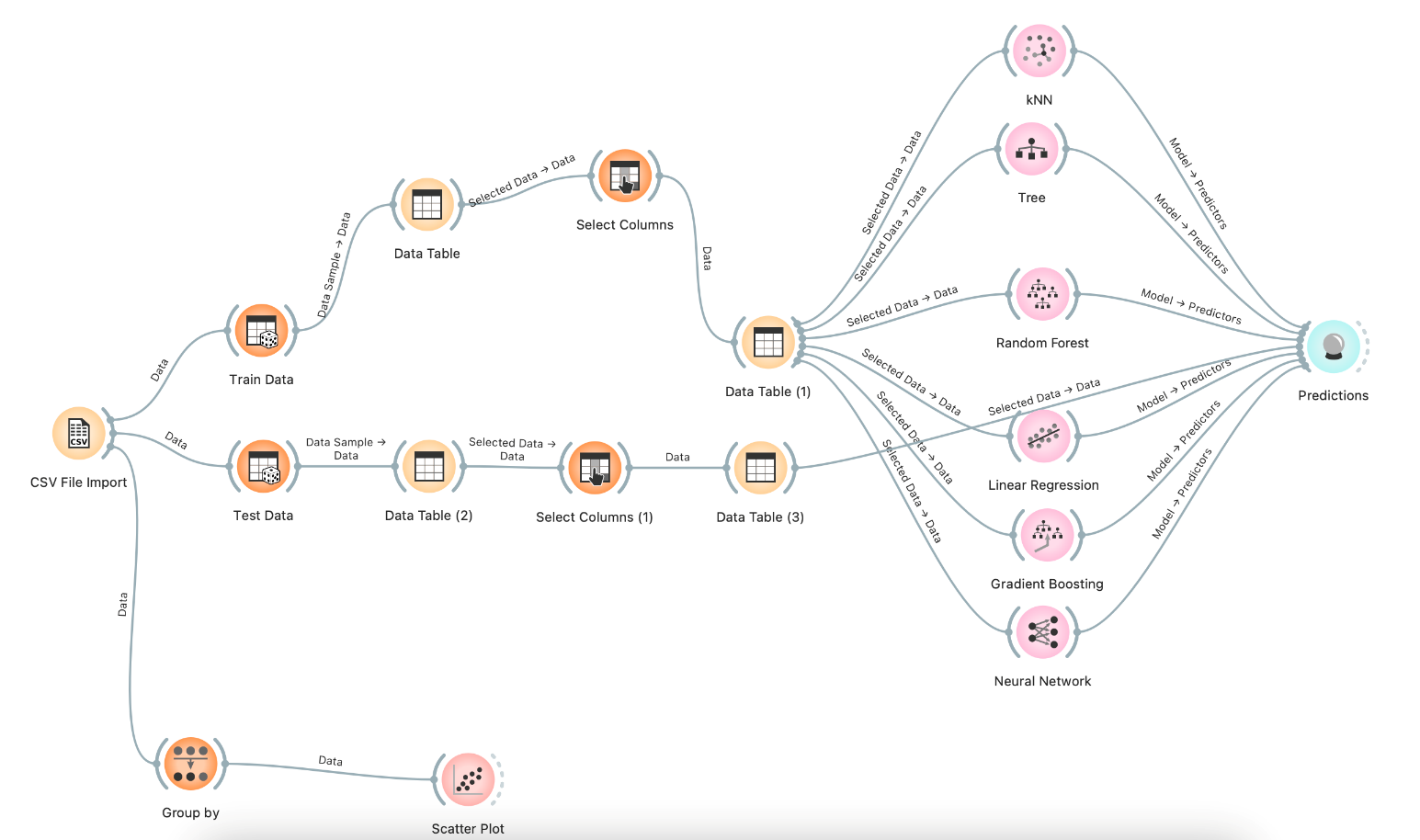


Figure 5 Orange Structure

As we can see from the Figure [5], we are loading our csv file into orange. Then we are partition our data into test and train with respect to the portion of 0.8. In the train data branch, we are selecting our predictors as “Gender”, “Demographics Response” and “Country” columns and our predicted value is “Value”. We are testing five different methods which are kNN, Tree, Random Forest, Linear Regression, Gradient Boosting and Neural Network.

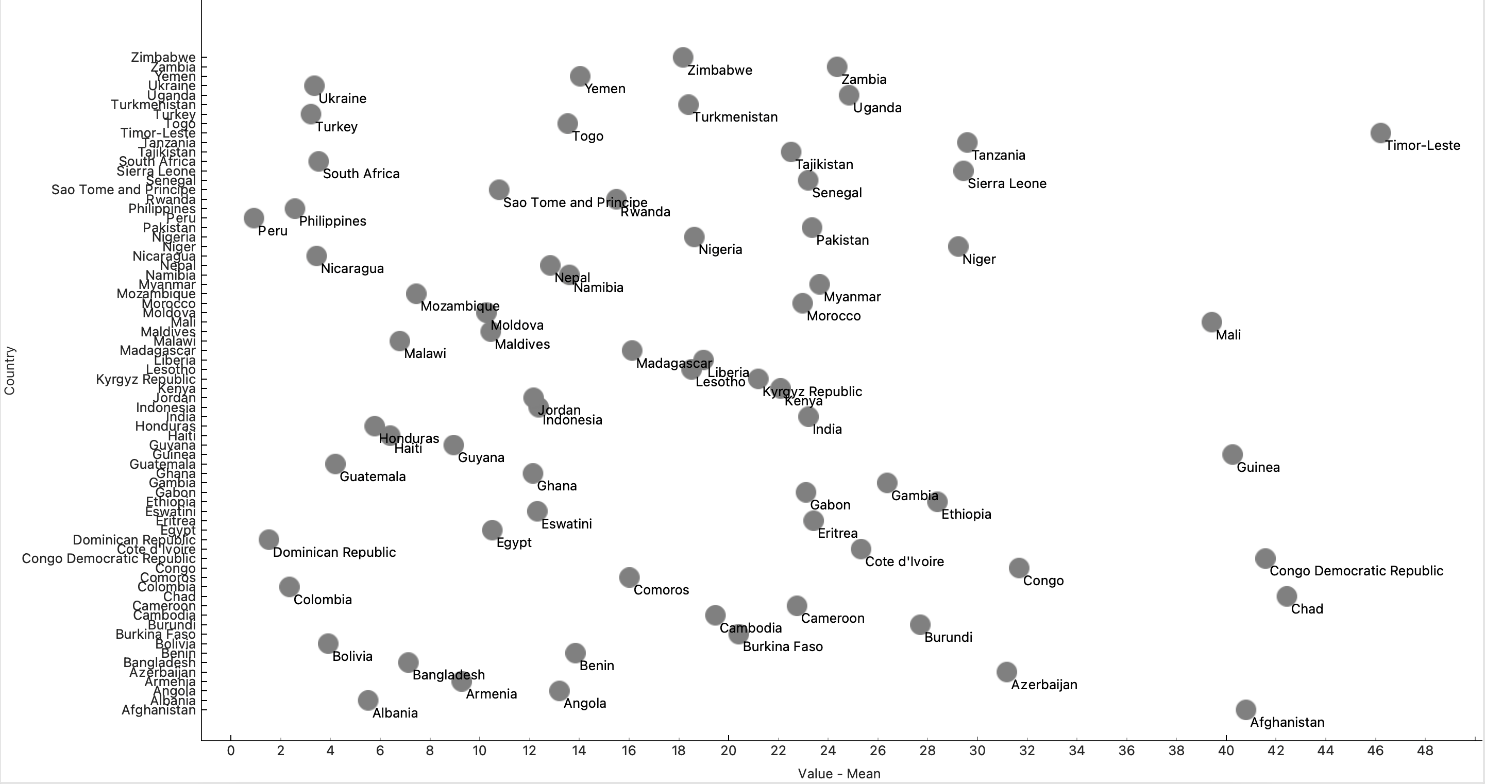


Figure 6 Country Value's

In different branch, we can see that we are grouping by “Country” column and you can observe the results in Figure [6]. Turkey has a value between 2-4 which corresponds to a good value actually. We will try to make a comparison between Turkey and around countries such as Ukraine, Bolivia, Philippines, Colombia etc.

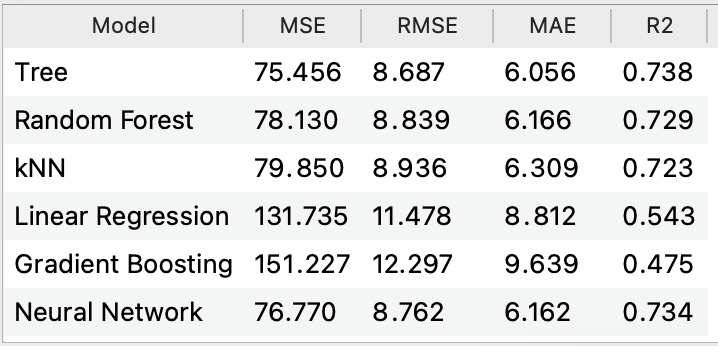


Figure 7 Mean Value Filled

From the Figure [7] we can observe that our best models are Tree, Random Forest, kNN and Neural Network. In all of those we have a R2 value with around 0.7-0.75 which is not good and not bad actually. Furthermore, our MSE, RMSE, MAE values are little bit higher than expected. The reason behind all of this is that, we don’t want to be in the overfit process because of that we did not use the “Question” column as a predictor. If we use it, value of R2 is approaching to 0.99 which is likely to be overfit.

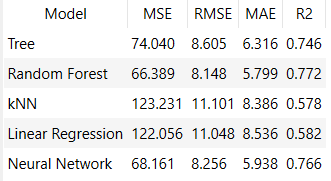


Figure 8 Deleted None Values

From the Figure [8], we are tried with deleting empty values. The reason behind that is, we had to round some values ​​regarding to only country because there was too much missing data. We thought that this could increase the error values ​​in our models. In this dataset, Tree, Random Forest and Neural Network still the bests but as you can see kNN values got worse when compare to Figure [7]. From the picture Random Forest is the best and we use it in our prediction process.

In the prediction process, as you can see from the Figure [6], we had split our data into test and train, we will use the 0.2 percentage of the dataset to find out our prediction values.

## 4.2-Prediction and Methods:

Decision Tree:

A decision tree is one of the supervised machine learning algorithms. This algorithm can be used for regression and classification problems — yet, is mostly used for classification problems. A decision tree follows a set of if-else conditions to visualize the data and classify it according to the conditions.

Random Forest:

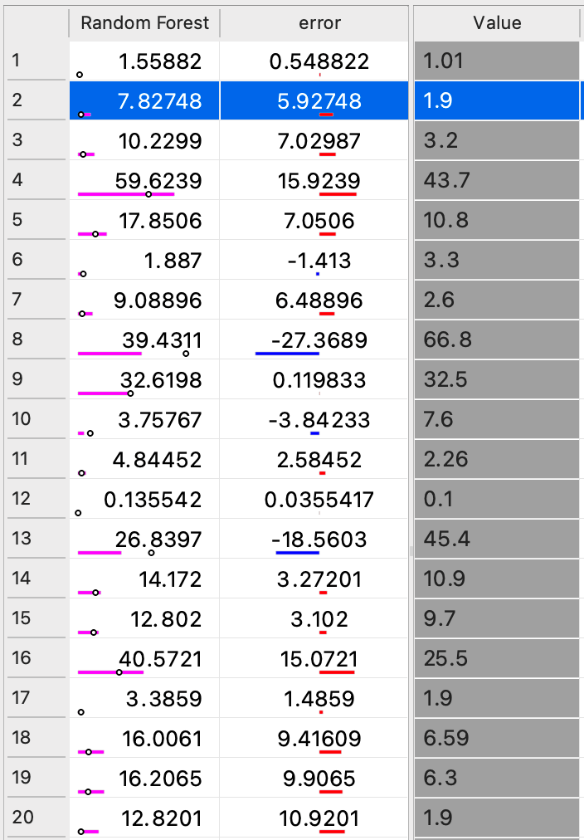
Random Forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. It builds decision trees on different samples and takes their majority vote for classification and average in case of regression.

kNN:

K-Nearest Neighbors Algorithm. The k-nearest neighbors’ algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. KNN works by finding the distances between a query and all the examples in the data, selecting the specified number examples (K) closest to the query, then votes for the most frequent label (in the case of classification) or averages the labels (in the case of regression).

Neural Network:

Neural networks are computing systems with interconnected nodes that work much like neurons in the human brain. Using algorithms, they can recognize hidden patterns and correlations in raw data, cluster and classify it, and – over time – continuously learn and improve.



Formun Üstü

Figure 9 Predicted and Actual Value

In the Figure [9], we can see that our predicted and actual values have same difference in each of the different partition. For example, in the first case our actual result is 1.01 but the predicted value is 1.55 and error is considerably small. But in the 13th example, predicted value is 26.83 but our actual value is 45.4. We can interpret that our model is not fed by the large Values because of that it gives us highly errored prediction. Another reason for that, for example, we think that in Afghanistan the rate will be much higher but, in some cases, their value might go down. Because of that machine predicts much higher or lower values.

As another approach to the Figure [9] we can say that our model’s predicted values considerably close to the actual values. We can find any country, gender and Demographics Response combination and find out the value of how much likely to your partner will physically or emotionally damage you.

Figure 10 Predicted vs Actual Values

From the Figure [10] we can see that, the predicted and actual value differs mostly in (3-5), (7-9) and (12-14) axis. In other cases, our model seems to have small portion of errors which indicates that it almost predicts the actual value. Other than that, our error value has kind of small with respect to another axis in the graph. Even though, random forest is the best model that we picked, it has large errors in some axis combinations actually.

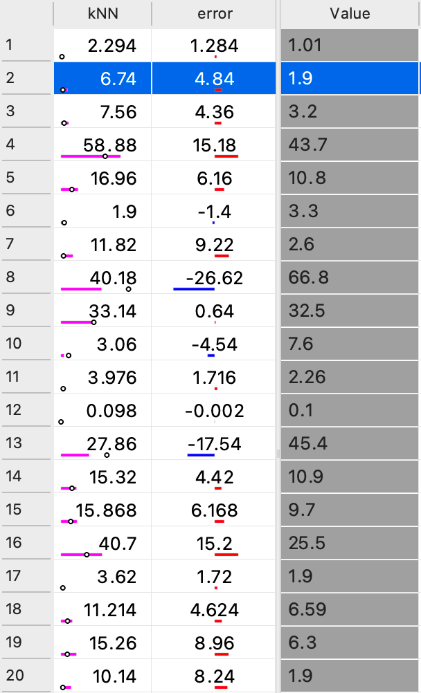


Figure 11 Predicted vs Actual kNN

Figure 12 Prediction and Actual Value (kNN)

From the Figure [11] and Figure [12] we can observe the kNN based model prediction values, actual values and error of these values. As we said from the previous section, we have similar errors in the same axis. For example, error value is considerably high in (3 - 6), (8-10) and (12-14). The main idea behind those errors might be the lack of predictor variable that we not used to not caught by overfit. Another reason might be the same reason Random Forest model have which is some countries have absurd different values between the questions and response of the values. Because of that, unique individuals might reduce the mean value of the country and that process lead to have high errors in the prediction process.

# 5. Constitutions of the Countries:

### Women’s Rights in The Turkish Constitution 1982 (rev. 2017)

* Minors, women and persons with physical or mental disabilities, shall enjoy special protection with regard to working conditions.
* Men and women have equal rights. The State has the obligation to ensure that thisequality exists in practice. Measures taken for this purpose shall not be interpreted ascontrary to the principle of equality.

### Women’s Rights In The Ukraine Constitution 1996 (rev. 2019)

* Equality of the rights of women and men is ensured: by providing women with opportunities equal to those of men, in public and political, and cultural activity, in obtaining education and in professional training, in work and its remuneration; by special measures for the protection of work and health of women; by establishing pension privileges, by creating conditions that allow women to combine work and motherhood; by legal protection, material and moral support of motherhood and childhood, including the provision of paid leaves and other privileges to pregnant women and mothers.
* The employment of women and minors for work that is hazardous to their health, is prohibited.

### Women's Rights In The Nicaragua Constitution [1987 (rev. 2014)](https://www.constituteproject.org/constitution/Nicaragua_2014?lang=en)

* Unconditional equality of all Nicaraguans in the enjoyment of their political rights, in the exercise of these rights, and in the fulfillment of their duties and responsibilities, is established; there exists absolute equality between men and women.
* The State grants special protection to the process of human reproduction. Women shall have special protection during pregnancy and shall be granted maternity leave with pay and all appropriate social security benefits. No one may deny employment to women for reasons of pregnancy nor dismiss them during pregnancy or the post-natal period; all in conformity with the law.

### Women's Rights In The Haiti Constitution[1987 (rev. 2012)](https://www.constituteproject.org/constitution/Malawi_2017?lang=en)

* To assure to women a representation in the instances of power and of decision which must conform to the equality of the sexes and to equity of gender.
* The principle of the quota of at least thirty percent (30%) of women is recognized at all levels of national life, notably in the public services.

### Women's Rights In The Egypt Constitution 2014 (rev. 2019)

* The state commits to achieving equality between women and men in all civil, political, economic, social, and cultural rights in accordance with the provisions of this Constitution.
* The state commits to the protection of women against all forms of violence, and ensures women empowerment to reconcile the duties of a woman toward her family and her work requirements.
* The House of Representatives is composed of no less than four hundred and fifty members elected by direct secret public ballot. At least one quarter of the seats shall be allocated to women.

When we look at the results of the above-mentioned constitutions and the results of the data we found, we observe that although Ukraine has a similar violence rate with Turkey, it emphasizes women's rights in detail in its constitution. In addition, when we look at average countries such as Haiti, we can see that the articles of women's rights in their constitutions have decreased. In other words, when we compare the violence rate results and the constitutions, if we consider that the data results are not unbalanced with Turkey's constitution, but these data are 5-10 years old from today's date, we can see that this rate may become unbalanced in the future due to the scarcity of women's rights articles in Turkey's constitution compared to other countries, and that the rate of violence will also increase. It should be noted that it can increase.

The results we found while interpreting the data; In addition to the countries mentioned above, by looking at the constitutions of 14 countries, we examined the constitutions of these countries in order to understand whether there is a relationship between the results we obtained and the constitutions, and moreover, if there is a country in a better position, why this country is in a better position.

References

Larxel (2020) Violence Against Women and Girls. Kaggle.

<https://www.kaggle.com/datasets/andrewmvd/violence-against-women-and-girls>

Worldometer. (2022). *Worldometer - real time world statistics*.

https://www.worldometers.info/

*Constitute*. (2022).

https://www.constituteproject.org/