Earthquake Prediction Model

Furkan BEKAR
Gazi University
furkanbekar.eng@gmail.com

ABSTRACT
In this article, we explain the formatting guidelines for the earthquake prediction model.

1. INTRODUCTION
Earthquake is the most hazards natural disaster seen in Turkey. Since the earthquake affects a wide area and causes a considerable loss of lives and property, it is the most important natural disaster in Turkey. 96 % of Turkey and 99 % of people living in Turkey are under the risk of earthquake .This makes the earthquake an important issue from social and economical point of view. Therefore, the reality of high earthquake risk should be taken into account in selecting residential and industrial areas, and in planning all other activities. As in the several countries, Turkey had big earthquakes in 1999. Especially in August 17, 1999 İzmit and November 12, 1999 Düzce earthquakes a lot of lives and property were lost, since the areas affected by these earthquakes are densely populated and have many industrial establishment. These earthquakes, have made people institutions and scientist to pay considérable attention to the earthquakes.

2. DATA SET
Our data set is associated with the earthquake in Turkey. We did not get our data set ready from somewhere. We created it with the data we received from the "KANDİLLİ OBSERVATORY AND EARTHQUAKE RESEARCH INSTITUTE". First, we received the data in txt format from Kandilli Observatory. Then, we opened the txt file we downloaded with Excell and converted it to csv file format. In the txt file, we only took the year part in the time part. There are values calculated with different formulas expressing the magnitude of the earthquake. however, there are rarely differences between these values. That is, almost all of them are the same value. some rows do not have some values calculated by formulas. We wrote a special script with python to collect the earthquake magnitude in a single column to avoid any corruption in the data set. We preferred to remove some unnecessary columns in the txt file we downloaded. Comma was used in float type values, and this was causing an error in the python language. That's why we wrote a function to replace commas with dots. After extracting the data, we removed the missing values from the data set by editing the missing values. We want to explain the columns in our data set. There are 5 different columns. First column is year column. The years in which the earthquakes occurred are shown here. Second column is latitude column. This column contains information about the latitude of earthquakes. Third column is longitude column. This column contains information about the longitude of the earthquakes. Fourth column is depth column. This column contains information about how deep earthquakes are. Fifth column is Mw column. This column contains information about the intensity of earthquakes. They are interconnected, as the destructiveness of earthquakes is linked to their intensity and depth. There are 17125 earthquakes in the data set. We gave a brief explanation about our data set. Let's move on to the analysis and visualization part.[1]

3. ANALYSIS
We made different analyzes according to years, latitude, longitude, depth and mw of earthquakes. We used many tables and graphics in the analysis part. We will show and explain them one by one below.

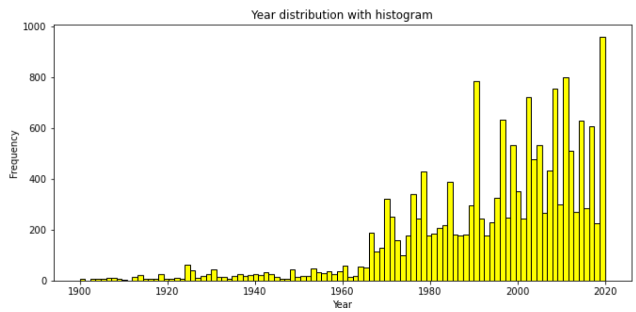

Figure 1. Year-Earthquake Frequency Histogram

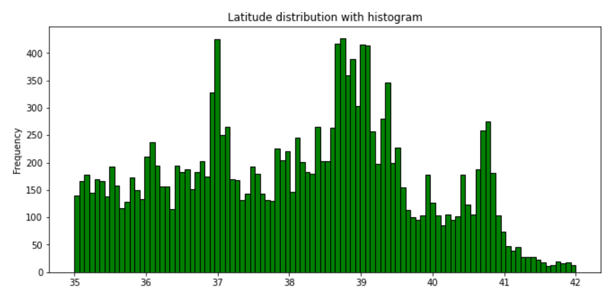

Figure 2. Latitude-Earthquake Frequency Histogram

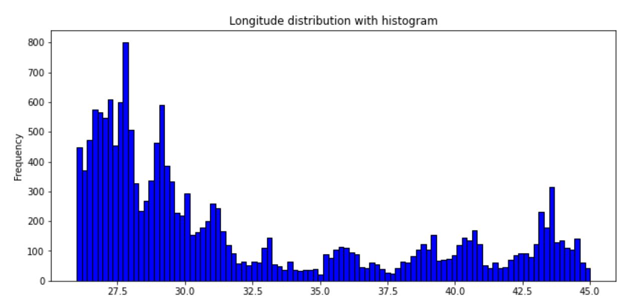

Figure 3. Longitude-Earthquake Frequency Histogram

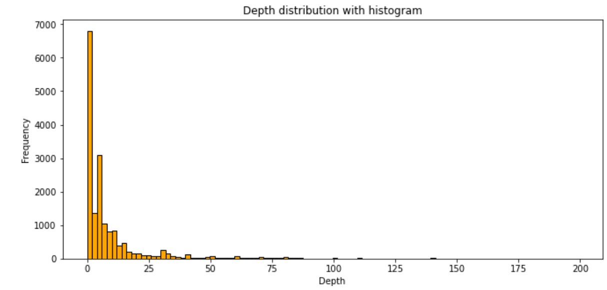

Figure 4. Depth-Earthquake Frequency Histogram

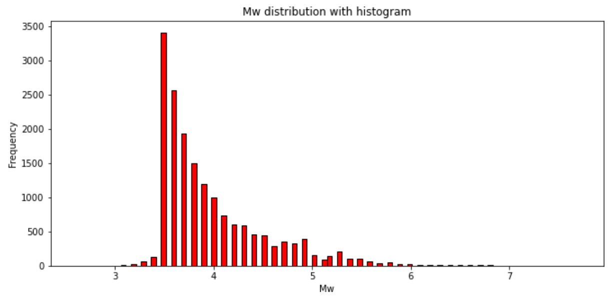

Figure 5. Mw-Earthquake Frequency Histogram

	Depth	Mw
389	57.1	6.100000
627	180.0	6.000000
586	130.0	5.633333
574	120.0	5.533333
441	70.9	5.500000
...	...	...
431	69.2	3.400000
339	40.9	3.300000
605	138.3	3.300000
527	92.4	3.200000
602	137.5	3.200000

629 rows × 2 columns

Table 1. Depth-Mw

Table 1 shows a Depth-Mw relationship.

	Mw	Depth
13	7.7	20.000000
12	7.5	1.000000
11	7.4	1.800000
10	7.2	35.833333
9	7.1	70.000000
8	6.9	16.000000
7	6.8	10.557143
6	6.7	20.083333
5	6.6	8.860000
4	6.5	31.875000
3	6.4	8.200000
2	6.3	25.855556
1	6.2	30.714286
0	6.1	32.772727

Table 2. Mw-Depth

Table 2 shows a Mw-Depth relationship.

	Mw	Latitude	Longitude
13	7.7	39.800000	39.510000
12	7.5	39.050000	44.040000
11	7.4	40.760000	29.970000
10	7.2	38.667067	33.983667
9	7.1	37.980000	44.480000
8	6.9	40.600000	27.200000
7	6.8	39.042443	33.156043
6	6.7	39.741667	33.975000
5	6.6	37.913860	31.885140
4	6.5	38.290275	34.074337
3	6.4	39.388000	31.598000
2	6.3	36.524978	31.017756
1	6.2	39.407857	30.811429
0	6.1	38.454600	34.168709

Table 3. Latitude Longitude and Size Relationship

Table 3 shows the relationship between latitude longitude and Mw. Thanks to these tables, we analyzed whether there is a relationship between mw, depth, latitude and longitude.

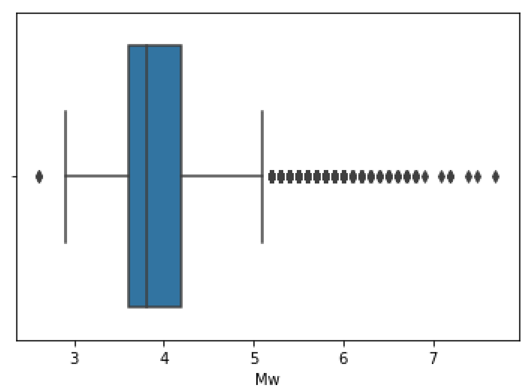

Figure 6. Mw Outlier

	Missing Value	Percent
Year	0	0.0
Latitude	0	0.0
Longitude	0	0.0
Depth	0	0.0
Mw	0	0.0

Table 4. Missing Value

There is no missing value because we eliminate the missing values.

4. VISUALIZATION

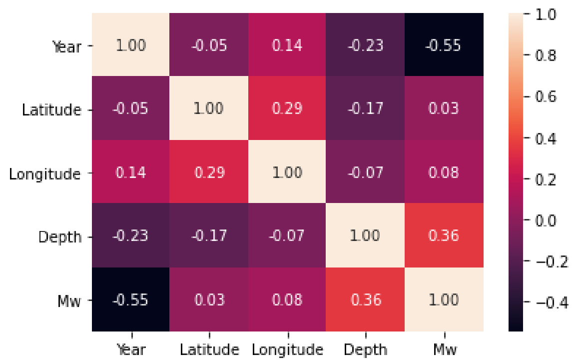

Figure 7. Correlation Between Year -- Latitude -- Longitude -- Depth – Mw

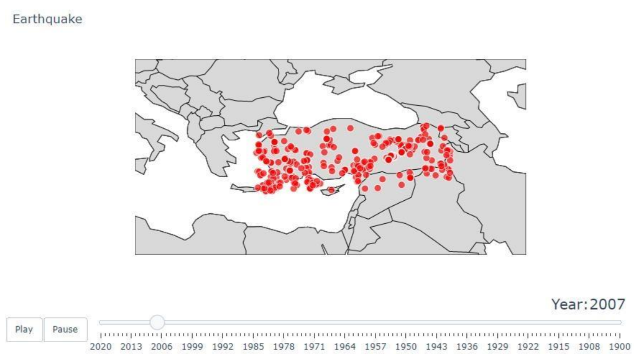

Figure 8. Earthquake Simulation by Years

The figure shows the earthquakes according to 8 years. You can see the earthquake in the year you want. According to the above simulation, we observe that there have been more earthquakes in recent years than in previous years.

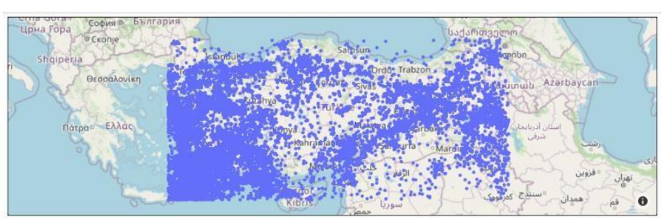

Figure 9. Locations of All Earthquakes

Figure 9 shows all earthquakes. When the locations of all past earthquakes are examined, it is observed that the most earthquakes are mostly on the Mediterranean and Aegean coasts.

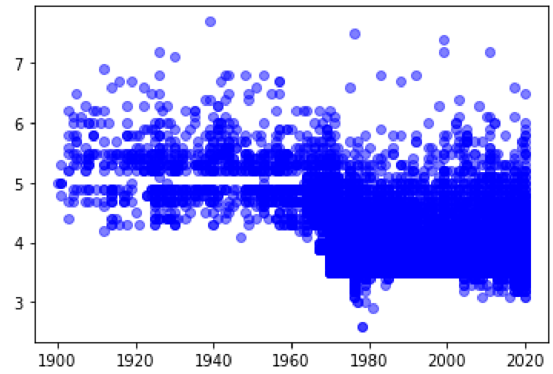

Figure 10. Year-Mw Relationship

It seems that there have been more earthquakes in figure 10 in recent years.

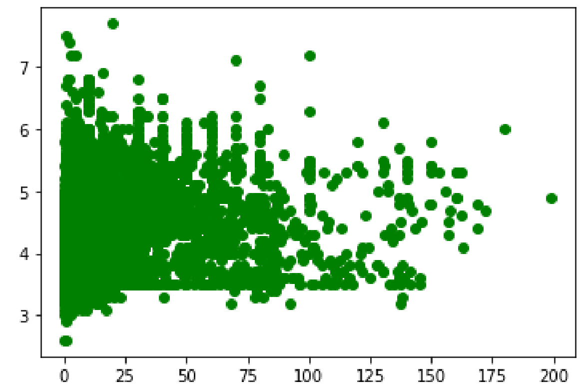

Figure 11. Depth-Mw Relationship

It is observed that there are mostly shallow earthquakes in our country. The shallowness of earthquakes causes an increase in the destructiveness of the earthquake. Unfortunately, our loss of life and property increases accordingly.

	Year	Latitude	Longitude	Depth	Mw	greater than six
0	2020	39.6762	27.8122	5.5	3.7	False
1	2020	35.5632	31.6060	5.0	3.6	False
2	2020	39.0075	35.8543	5.0	3.4	False
3	2020	38.1787	38.6887	8.2	4.1	False
4	2020	39.0807	27.6843	7.5	3.5	False
...	...	...	...	...	...	...
17120	1901	37.8000	27.8000	15.0	5.0	False
17121	1901	38.4000	31.4000	5.0	5.0	False
17122	1901	38.2000	27.7000	5.0	5.0	False
17123	1901	37.9000	27.9000	15.0	4.8	False
17124	1900	37.8000	29.1000	5.0	5.0	False

17125 rows × 6 columns

Table 5. Greater than Six Mw Earthquakes

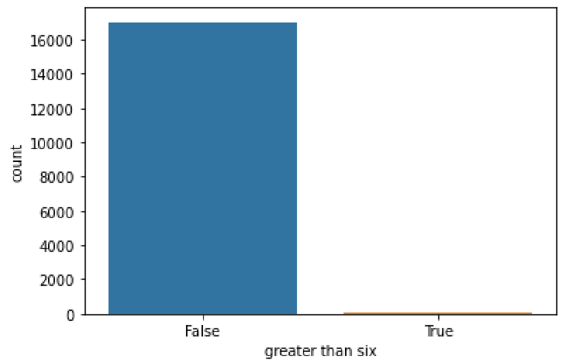

Figure 12. Greater than Six Mw Earthquakes

If we examine figure 12 and table 5 together, we reach the following conclusion. There are many earthquakes daily in our country, but earthquakes with high destructive effects occur at regular intervals.

	Year	Latitude	Longitude	Depth	Mw	greater than six	Class
0	2020	39.6762	27.8122	5.5	3.7	False	shallow
1	2020	35.5632	31.6060	5.0	3.6	False	shallow
2	2020	39.0075	35.8543	5.0	3.4	False	shallow
3	2020	38.1787	38.6887	8.2	4.1	False	shallow
4	2020	39.0807	27.6843	7.5	3.5	False	shallow
...	...	...	...	...	...	...	...
17120	1901	37.8000	27.8000	15.0	5.0	False	shallow
17121	1901	38.4000	31.4000	5.0	5.0	False	shallow
17122	1901	38.2000	27.7000	5.0	5.0	False	shallow
17123	1901	37.9000	27.9000	15.0	4.8	False	shallow
17124	1900	37.8000	29.1000	5.0	5.0	False	shallow

17125 rows × 7 columns

Table 6. Earthquake Class

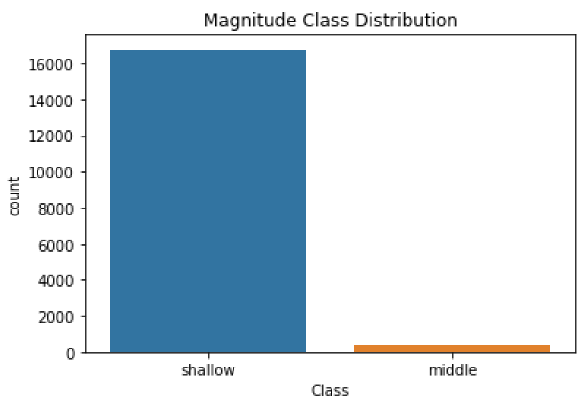

Figure 13. Earthquake Class

If we examine figure 13 and table 6 together, we reach the following conclusion. Shallow earthquakes are generally observed in our country. This result tells us that our country is geographically young.

	Year	Latitude	Longitude	Depth	Mw	greater than six	Class	destructiveness
0	2020	39.6762	27.8122	5.5	3.7	False	shallow	no significant damage occurs
1	2020	35.5632	31.6060	5.0	3.6	False	shallow	no significant damage occurs
2	2020	39.0075	35.8543	5.0	3.4	False	shallow	no significant damage occurs
3	2020	38.1787	38.6887	8.2	4.1	False	shallow	no significant damage occurs
4	2020	39.0807	27.6843	7.5	3.5	False	shallow	no significant damage occurs
...	...	...	...	...	...	...	...	...
17120	1901	37.8000	27.8000	15.0	5.0	False	shallow	Damage may occur in structures close to the ea...
17121	1901	38.4000	31.4000	5.0	5.0	False	shallow	Damage may occur in structures close to the ea...
17122	1901	38.2000	27.7000	5.0	5.0	False	shallow	Damage may occur in structures close to the ea...
17123	1901	37.9000	27.9000	15.0	4.8	False	shallow	no significant damage occurs
17124	1900	37.8000	29.1000	5.0	5.0	False	shallow	Damage may occur in structures close to the ea...

17125 rows × 8 columns

Table 7. Destructive Earthquakes

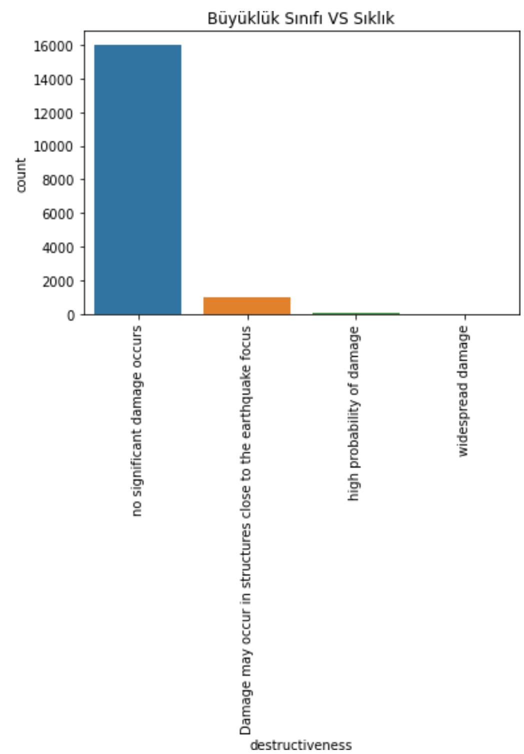

Figure 14. Magnitude Class Distribution

If we examine figure 14 and table 7 together, we reach the following conclusion. Although most of the earthquakes that occur do not seem to cause damage, the number of major earthquakes in our country is considerably high.

	magnitude level	Mw
0	low	3.7
1	low	3.6
2	low	3.4
3	hight	4.1
4	low	3.5
5	low	3.4
6	hight	4.0
7	low	3.5
8	low	3.8
9	low	3.6
10	hight	4.4

Table 8. Earthquake - Magnitude Level

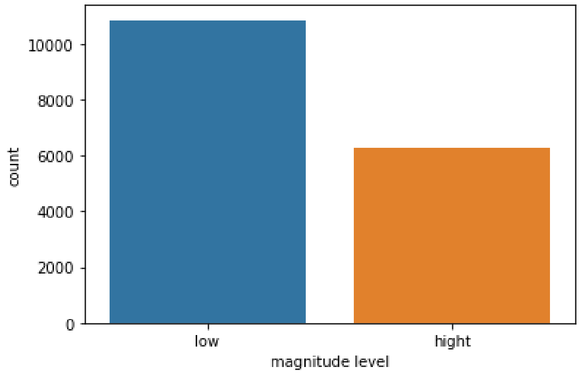

Figure 15. Earthquake - Magnitude Level

If we examine figure 15 and table 8 together, we reach the following conclusion. The table shows us that the rate of large earthquakes is very high. Therefore, prevent it as soon as possible to minimize loss of life and property.

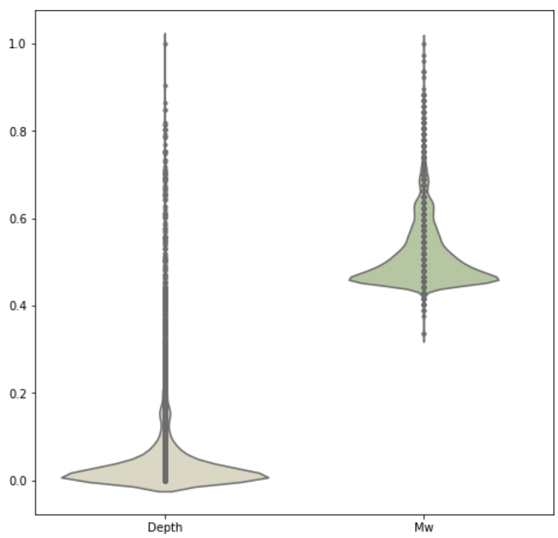

Figure 16. All Value Densities of the Scales

5. MODELS

We created 4 different models to find the best model.

1. Decision Tree Classifier
2. K Neighbors Classifier
3. Gaussian NB
4. Logistic Regression

5.1 Decision Tree Classifier

```
DecisionTreeClassifier(criterion='entropy', max_depth=4)
for decision tree accuracy:  1.0
[1. 1. 1. 1. 1. 1. 1. 1. 1. 1.]
cv_scores mean:1.0
```

	precision	recall	f1-score	support
Damage may occur in structures close to the earthquake focus	1.00	1.00	1.00	309
high probability of damage	1.00	1.00	1.00	16
no significant damage occurs	1.00	1.00	1.00	4812
widespread damage	1.00	1.00	1.00	1
accuracy			1.00	5138
macro avg	1.00	1.00	1.00	5138
weighted avg	1.00	1.00	1.00	5138

Figure 17. Decision Tree Classifier Model

5.2 K Neighbors Classifier

```
for knn accuracy 0.976449980537174
[0.94045534 0.94454174 0.93578517 0.94921191 0.95680093 0.95560748
 0.96845794 0.96028037 0.93341121 0.29556075]
cv_scores mean:0.8840112853331586
```

	precision	recall	f1-score	support
Damage may occur in structures close to the earthquake focus	0.72	0.87	0.79	254
high probability of damage	0.38	0.67	0.48	9
no significant damage occurs	1.00	0.98	0.99	4875
widespread damage	0.00	0.00	0.00	0
accuracy			0.98	5138
macro avg	0.52	0.63	0.56	5138
weighted avg	0.98	0.98	0.98	5138

Figure 18. K Neighbors Classifier Model

5.3 Gaussian NB

```
for gaussian naive base accuracy 0.9614636045153756
[0.97022767 0.97431407 0.97606538 0.99766492 1.         0.99941589
 1.         1.         0.99824766 0.47546729]
cv_scores mean:0.9391402878482851
```

	precision	recall	f1-score	support
Damage may occur in structures close to the earthquake focus	0.92	0.62	0.74	457
high probability of damage	1.00	1.00	1.00	16
no significant damage occurs	0.96	0.99	0.98	4664
widespread damage	1.00	1.00	1.00	1
accuracy			0.96	5138
macro avg	0.97	0.90	0.93	5138
weighted avg	0.96	0.96	0.96	5138

Figure 19. Gaussian NB Model

5.4 Logistic Regression

```
for logistic regression accuracy 0.9560140132347217
[0.93636894 0.93636894 0.9392878  0.94220665 0.95271454 0.97371495
 0.9953271  0.99591121 0.99591121 0.375     ]
cv_scores mean:0.9042811362805594
```

	precision	recall	f1-score	support
Damage may occur in structures close to the earthquake focus	0.52	0.68	0.59	238
high probability of damage	0.00	0.00	0.00	0
no significant damage occurs	0.99	0.97	0.98	4900
widespread damage	0.00	0.00	0.00	0
accuracy			0.96	5138
macro avg	0.38	0.41	0.39	5138
weighted avg	0.97	0.96	0.96	5138

Figure 20. Logistic Regression Model

It has an accuracy of over 95% in the 4 models we created. This means that it is successful in 4 models. We will choose the most successful among us. The best model for us is the decision tree. The decision tree gave the best result.

6. CONCLUSION

Earthquake is a big problem for our country. There is loss of life and property as a result of the earthquake. We created the Earthquake Prediction Model to avoid these. With this model, we can predict major earthquakes that may occur. Here we selected the most successful one using 4 different models. The most successful model is the decision tree model.

7. REFERENCES

[1] "BOĞAZİÇİ ÜNİVERSİTESİ KANDİLLİ RASATHANESİ VE DEPREM ARAŞTIRMA ENSTİTÜSÜ (KRDAE) BÖLGESEL DEPREM-TSUNAMİ İZLEME VE DEĞERLENDİRME MERKEZİ (BDTİM)", URL:http://www.koeri.boun.edu.tr/scripts/lst0.asp