# Failure Prediction with Construction Machinery Oil Analysis Results

**Dr.Semra Erpolat Taşabat[1], Tayfun Özçay[2], Esra Akca[3], Nermin Yalçı[4], Canberk Arıcı[5]**

*[1]Mimar Sinan Fine Arts University, semra.erpolat@msgsu.edu.tr*
*[2,3,4,5]Borusan Makina ve Güç Sistemleri Sanayi A.Ş*

With the advancement of technology and widespread use of the internet, the concept of big data, which we often hear about, has emerged. Transforming big data, briefly defined as a stack of unstructured data, into meaningful information and revealing hidden patterns can be achieved by different methods. The use of this group of methods and algorithms, called data mining, with artificial intelligence and statistics, enables more understandable, meaningful and effective decisions to be made. This helps to optimize profits by reducing costs and increasing performance.
In this study; By Borusan Cat, Caterpillar Inc. Using oil analysis data obtained from construction machines, a model has been developed that will enable early malfunction detection of machines and identification of vehicle maintenance needs. With the developed system, the oil analysis data of the vehicles are used to predict the number of working hours in which the vehicle will break down, using machine learning methods, in advance. Then, forecast data is integrated into decision mechanisms in business processes, and finally, the information obtained is reported using data visualization technologies and made traceable through summary data.
The system developed in this personalized product-focused study, which is the subject of Industry 4.0, can be easily adapted to the operation of different machines. In this way, it will be easier to track and control the vehicles, and it will be possible to detect any malfunctions that may occur in the vehicles without stopping the flow of the process. By contributing to the extension of the life of the machines used with the proposed approach, eliminating the costs of purchasing a new machine, which can be quite costly, or purchasing repairs and spare parts due to possible damage, will provide significant returns to companies in terms of cost and time. Another important contribution of the system will undoubtedly be environmental sustainability.

**Keywords:** Machine Failure, Machine Learning, Decision Tree Algorithm, Oil analysis, Environmental Sustainability.

## 1. Introduction

Industry 4.0 emerges as a new vision with developing technology. The ability to intervene in the system before a malfunction occurs, high efficiency, reduction in production costs and product time to market, and the flexibility it provides to production have made Industry 4.0 advantageous. As one of the personalized services with Industry 4.0, we can exemplify the practice of preventing disruptions by predicting machine malfunctions in advance. According to the result of this, technical teams; Instead of "repairing what is broken", were able to focus on their main job, "maintenance to prevent malfunctions". However, the first function of eliminating the main sources of failure is not to change the task structure of maintenance teams but mainly to significantly increase machine and component life while reducing maintenance costs. Oil maintenance lies at the basis of the maintenance technology that was put forward in the 1980s as Proactive Maintenance (Preventive Maintenance). Studies have shown that the source of 42% of mechanical failures is oil contamination and lubrication-related problems. Therefore, conscious use of oil and protection of oil quality are important elements. For this reason, oil analyses are very important for those who use the machines. While oil analyses are important, knowing what the values in the analysis report mean and making sense of them are much more important than analysis. Oil analysis is a preventive maintenance program used and implemented to increase the life and efficiency of work machines and to prevent maintenance costs and time loss (James et al., 2017). All used oils through oil analysis; Engine, transmission, differential, traction, and hydraulic system oils can be tested. These tests were developed to evaluate not only the condition of the oil but also the condition of work machines and engines. The success of a preventive maintenance program can be measured by the financial gain it brings to the user. The application of oil analysis is cheap, quick, and easy, and it can also provide significant savings. Oil analysis application, combined with the high-performance properties of the original oil and filters used, forms the basis of the best possible maintenance program for your machines.

This study, which comes into contact with the principles on which Industry 4.0 is based, was created by the integration of mechanical processes and information technologies. The study generally covers the steps of real-time monitoring of analysis outputs and how many hours it will take before a possible malfunction occurs. For this purpose, Caterpillar Inc. is sold by Borusan Cat, one of the world's well-known companies in the field of production and maintenance of construction equipment. Periodic oil analysis results from the machines were used to detect major malfunctions that would require part replacement in construction vehicles. Particulate data regarding the oil contained in construction machines is provided by the laboratory results of the samples taken. It is decided whether the oil can continue to serve or not after the characteristics of the oil are defined through analysis.

The traditional approach in general; By interpreting the data obtained as a result of oil analysis according to certain value ranges; aims to increase the life and efficiency of the machines and to prevent loss of time by reducing maintenance and repair costs, and it only fulfills the function of fault detection. Detecting engine and transmission failure by looking at the particles in the oil sample can be given as an example of the working principle of the traditional approach. The proposed approach is aimed not only to evaluate the data obtained from the oil samples taken from the machines at certain periods according to the reference oil value ranges but also to include the data collected from all processes up to the oil analysis change in the machines and to process them with machine learning methods and make a fault detection prediction. In this way, the degree to which the machine is approaching failure will be estimated by using the particle values in each oil sample taken, regardless of the value ranges previously determined for each machine.

In this way, construction machines will be able to contribute to the field for a longer period and their environmental sensitivity will also increase. Namely: While metal elements mixed with oil can damage different machine components and cause them to deteriorate, on the other hand, burning oils with metal mixtures increases the damage to the environment and can cause negative effects. In this way, benefits are provided for the environment, economy, and society, which are the three basic components of sustainability.

## 2. LITERATURE REVIEW

Although there is no similar study in the literature, it is possible to come across different studies conducted in this field. One of these; is the article titled Oil Analysis and Importance of Construction Equipment, published in the TMMOB Chamber of Mechanical Engineers Construction Equipment Symposium and Exhibition in 2003. Detailed particle examinations regarding oil analysis and the determination of damages that will occur as a result of oil analysis are mentioned. Details that play a role in maintaining oil quality and extending machine life are listed (İbrahim, 2003).

In a different research, the issues of regulating the usage time of lubricating oil samples used in marine-type diesel engines and thus reducing maintenance costs were examined experimentally. Regular samples are taken from the marine diesel engine at certain intervals and the kinematic viscosities, densities, sediment, particle number, water, flash point, pour point, ICP (Inductively Coupled Plasma) element analysis, and ASTM (American Society for Testing and Materials) standards of these samples are determined. was determined accordingly (Burak,2007).

In an article written by Wakiru, James; It was studied to separate the pollution in the oil using the hierarchical clustering analysis method. The study aims to obtain output for making maintenance decisions for the engines. As a result of the study, there is uncertainty in the results due to sampling errors (James et al., 2018).

Unlike the use of oil analysis results on fault prediction from work machines in the literature, there are studies on fault prediction using IoT data. Taşabat et all. In the article written, there is a way to predict the malfunction of the machine by taking the IoT data on the work machine instantly (Semra et al., 2020). The progress procedure in the work processes of the predicted machines is similar to the action plan in the current study.

## 3. DATA SUMMARY

The number of machines that Borusan Cat sells and maintains is growing day by day. The data obtained from machines is also increasing rapidly at this rate. Transforming the available data piles into meaningful information and making them usable in making useful decisions is only possible with scientific analysis techniques. In this process, cleaning, organizing, modeling, and interpreting the data constitute the basic steps.

S.O.S programmed oil analysis performed with the S.O.S application used within Borusan Cat performs the function of testing used oil samples with advanced methods within 24 hours. Used oil samples taken from the machine are tested with advanced methods, allowing important determinations regarding both the malfunction and wear of the machine and the quality of the oil (https://www.borusancat.com/tr/service/sos-oil-analysis).

By combining this traditionally used system with the machine learning approach proposed in the study, fault detection predictions can be made for the machines analyzed during their future lifetimes. In this way, it was possible to achieve the following advantages:

• Problems can be diagnosed early, so small problems can be repaired before they turn into major malfunctions.

• Ensures effective budget management by estimating revision time and machine life.

• It detects positive as well as negative situations and prevents expenses for unworn components.

• Confirms the suitability of oil change intervals.

• Predicting when the malfunction may occur

With the proposed system, analysis and predictions can be made not only for Cat machines and Cat oils but also for other brands of machines and oils.

Figure 1 shows the steps of the traditional approach and the proposed new approach in evaluating oil analysis results.

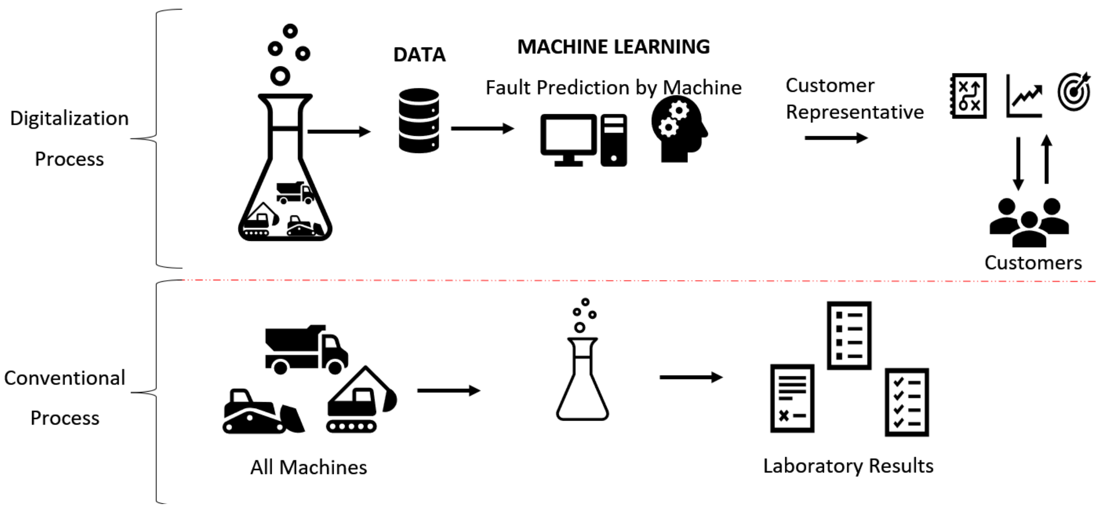


Figure 1: Traditional process-Digital process

As can be seen in Figure 1; in the traditional process, all machines come to the service oil analysis is performed and the results are shared with customers. In the proposed new approach, the oil analysis results of the machines that come to the service and have their oil analyzed with a digital process based on artificial intelligence are modeled with machine learning methods, and predictions are made about when the machines will malfunction.

## 4. OBTAINING DATA

To monitor and increase the efficiency of work machines, it is important to take oil analysis samples at certain periods and check the analysis results. By storing the analysis results performed at regular intervals within Borusan Cat, machine oil analysis data is obtained by utilizing cloud computing. Oil analysis data for 240 machines operating in different areas and regions in the field are recorded in the data warehouse in the format given in Table I.

Table I: Slice of raw data

| Sample Date | Machine | Work Hours | Particle Type | Particle Name | Value |
|---|---|---|---|---|---|
| 2019-19-07 | Machine 1 | 8691 | Metals | Mo | 3 |
| 2019-17-01 | Machine 2 | 10377 | Metals | Cu | 3 |
| 2019-11-01 | Machine 3 | 7185 | Metals | Mg | 383 |
| 2018-03-12 | Machine 4 | 1726 | Metals | Al | 0 |
| 2019-01-04 | Machine 5 | 19025 | Metals | Mo | 0 |
| 2018-04-12 | Machine 6 | 2031 | Metals | Sn | 0 |
| 2018-27-11 | Machine 7 | 2143 | Metals | Fe | 45 |
| 2018-13-08 | Machine 8 | 16858 | Metals | Zn | 1485 |
| 2019-31-05 | Machine 9 | 2660 | Metals | Al | 3 |
| 2018-11-06 | Machine 10 | 8074 | Metals | Ca | 3998 |

Abbreviations for particle symbols in the stored data are given in Table II (Eric,2011).

Table II: Particle symbols explanations

| Elements | Symbol |
|---|---|
| Aluminum | Al |
| Boron | B |
| Chromium | Cr |
| Copper | Cu |
| Iron | Fe |
| Bullet | Pb |
| Magnesium | Mg |
| Mobile | Mo |
| Nickel | Ni |
| Silicon | Si |
| Silver | Ag |
| Sodium | Na |

| | |
|---|---|
| Tin | Sn |
| Titanium | Ti |
| Zinc | Zn |

Within the scope of the study, the particles listed above were considered independent variables, while the machine failure hour interval, which is given in Table III and consists of three categories, was considered the dependent variable. After receiving the particle results, predictions of the path to failure are made according to the groups of the dependent variable. Historical data of each machine where oil analysis is performed is recorded. In this way, particle changes of the relevant machine over time can be seen. Since the estimates will be calculated based on machine operating hours, the data format is also arranged according to machine operating hours. The model machine is operated according to working hours.

Table III: Target work hours

| Target Category | Target Category Exp. |
|---|---|
| C | 0-400 Difference work hour |
| A | 400-800 Difference work hour |
| T | 800-1000 Difference work hour |

## 5. DATA PREPROCESSING

The raw data obtained from the samples taken must first be preprocessed. For this purpose, procedures such as Complementing Missing Data, Removing Duplicate Data, Transformation, Integration, Cleaning, Normalization, and Dimension Reduction were performed on the raw data:

• In addition to statistical analysis, business information and logical cleaning operations were carried out and records of the same date and from the same machine were cleaned in the data set.

• The relevant machine data, which shows a decrease in working hours instead of increasing or remaining constant with the date on the data accumulated on a date basis, has been cleared.

• For the work machines whose oil samples were taken before going to the service, the working hours of going to the service and the working hours when the oil sample was taken were obtained from the database, and the data of the relevant machines were collected in a single field and a data set was created.

• By calculating the difference in working hours between taking the oil sample and going to the service, it was determined how many hours after the sample was taken and the time it took for the work machine to go to the service was determined, and target flagging was carried out in how many hours it took for the malfunction to occur.

• Transpose variables were derived from the particle names in the samples. As seen in Table IV, while the particle names in the laboratory results were kept one under the other as in the name field, the data set was extended to the right with the transpose process in the format seen in Table V.

Table IV shows the format of the raw data mentioned above. Table V shows an image of the corrected data obtained as a result of data pre-processing procedures.

Table IV: Data table

| Machine | Component | Value | Particles | Date |
|---|---|---|---|---|
| Machine 1 | Engine | 42 | Fe | 2018-01-08 |
| Machine 1 | Engine | 8 | Zn | 2018-01-08 |
| Machine 1 | Engine | 3 | Na | 2018-01-08 |

Table V: Transposed data

| Machine | Component | Fe | Zn | Na |
|---|---|---|---|---|
| Machine 1 | Engine | 42 | 8 | 3 |

## 6. METHODOLOGY

The data obtained within the scope of the study was cleaned, sorted, arranged, and formatted suitable for the model and tested with the decision tree algorithm, one of the machine learning methods. The reason for choosing a decision tree as a method is to create reference values by creating rules for particle values. In this way, early fault detection was possible for the machines examined.

In other words, in the study, it was estimated how many hours after the oil sample was taken the fault would appear in the target category variables Figure 3 with the decision tree modeling library of the Python programming language. Details about the developed approach are given in the subheadings below.

## 7. DECISION TREE ALGORITHM AND EVALUATION METRIC

Inductive learning algorithms that use the divide-and-conquer approach display information in a decision tree structure (Akgöbek and Öztemel, 2006). Decision trees are one of the tree-based algorithms used in classification and regression problems and can be used in complex data sets (James et al., 2018). Figure 2 shows the decision tree structure.

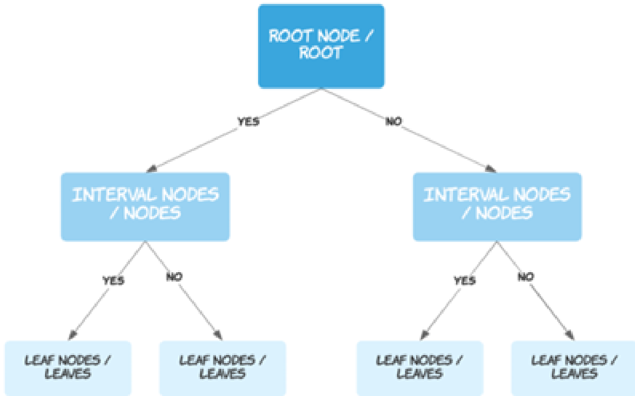


Figure 2: Decision Tree

The first cells of decision trees are called roots (root or root node). Each observation is classified as "Yes" or "No" based on the root condition. There are nodes (interval nodes) under the stem cells. Each observation is classified with the help of nodes. As the number of nodes increases, the complexity of the model increases. There are leaves (leaf nodes or leaves) at the bottom of the decision tree. The leaves provide the results.

The purity level of each subset as a class is obtained by the Gini coefficient (Muhammad et al.,2019). The Gini coefficient is given in Equation (1).

$$Gini = 1 - \sum_j p_j^2 \qquad (1)$$

Pj in the formula expresses the probability of occurrence of class j. This value is calculated for each class and the sum of the squares of the results is subtracted from one. The Gini value takes a result between 0 and 1, and the closer the result is to 0, the better the discrimination is made.

## 8. ALGORITHM CALCULATION

The code in Equation 1 was applied to the compiled data set in the Python programming language. The result of the resulting decision tree is shown in Figure 3.

```
#from matplotlib import pyplot as plt
from sklearn import datasets
from sklearn.tree import DecisionTreeClassifier
from sklearn import tree
```

```
clf = DecisionTreeClassifier(random_state=1234, min_samples_leaf=10)
model = clf.fit(all_t[coluse], all_t[target])
```

```
import graphviz
# DOT data
dot_data = tree.export_graphviz(clf, out_file=None,
                                feature_names=coluse,
                                class_names=target,
                                filled=True)

# Draw graph
graph = graphviz.Source(dot_data, format="png")
graph
```

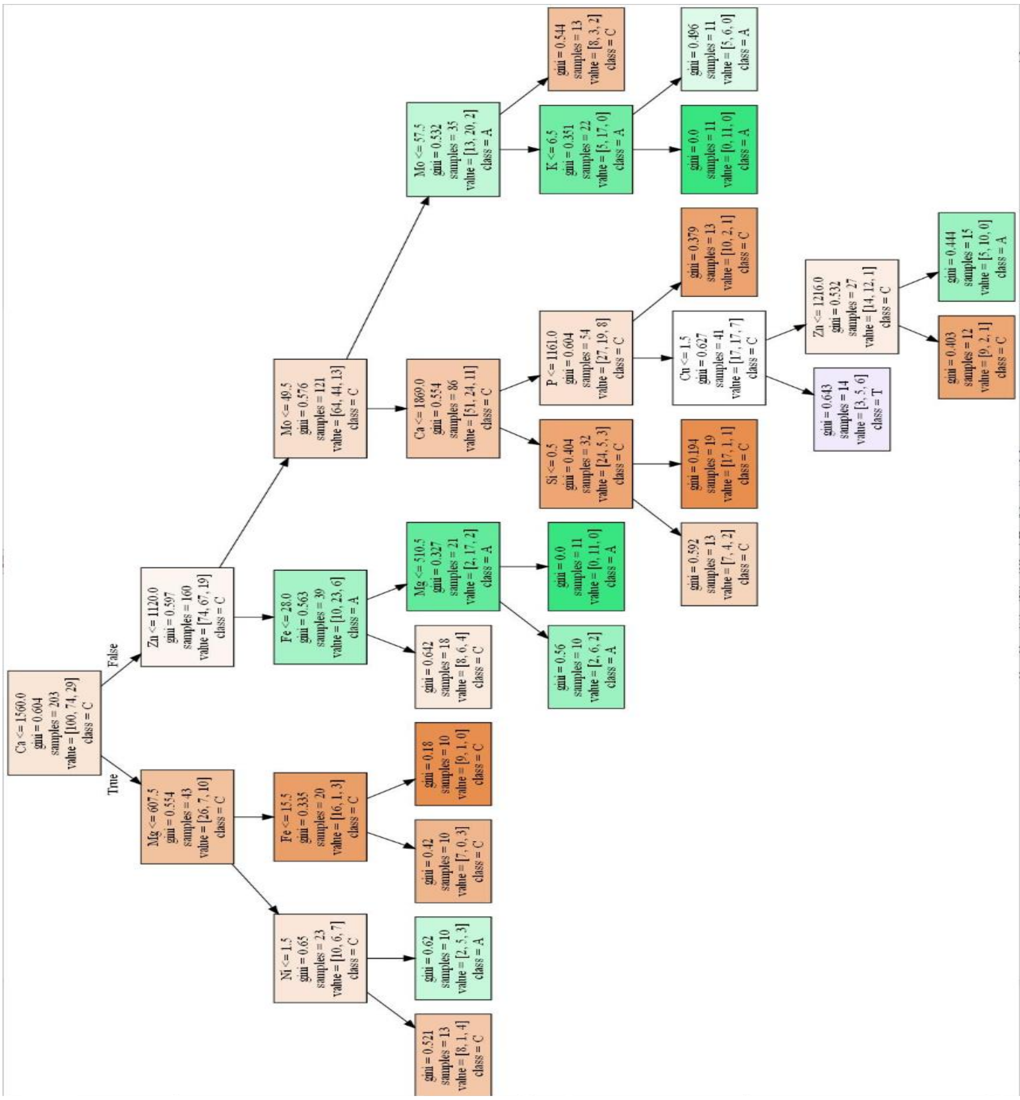

Figure 3: Decision Tree Algorithm result

## 9. CONCLUSION

Data mining, artificial intelligence, and statistics are concepts that include important points about the correct, effective, purposeful use of data and what to do with the information, rather than the size of the data.

In general terms, the main purpose of machine learning applications is; to learn patterns from data and create value using these patterns. For this purpose, a decision support system application was implemented by revealing valuable, usable information from big data. With the proposed approach, the function of predicting machines that are likely to need maintenance in the future has been achieved by using the decision tree model results developed with the oil data obtained from the machines under maintenance.

In the proposed approach, particle values approaching 0 with the Gini coefficient were examined as a model review metric using the decision tree algorithm. The findings obtained are listed below.

After the oil analysis:

• For C, which is defined as the target category that shows a tendency to malfunction in the 0-400 hour period, the Ca particle value must be less than 1560, the Mg value must be less than 607.5, and the Fe value must be greater than 15.5.

• For A, defined as the target category that tends to fail within 400-800 hours; the Ca value must be greater than 1560, the Zn value must be greater than 1120, the Mo value must be greater than 49.5, and the K value must be less than 6.5.

• For T, defined as the target category that tends to fail within 800-1000 hours; the Ca value must be greater than 1560, the Zn value must be greater than 1120, the Mo value must be less than 49.5, the Ca value must be greater than 1869, P value must be less than 1161, Cu value must be less than 1.5.

As can be seen from the results, it is confirmed by the proposed model that as the working hours after sampling increase, the number and diversity of particles also increase.

Since it was desired to obtain rule-based results within the scope of the study, the data set created using the decision tree algorithm, one of the machine learning techniques, was tested. As a result of the analysis, the modal success rate is 71%.

To apply the decision tree rules obtained with the proposed model on existing work machines, the model will be integrated into the company's system in the first place. In this way, information will be provided to the technical teams dealing with the controls of the work machines that are predicted to malfunction. As a result of this information, the technical team will reach the work machines and provide the necessary fault maintenance, control, and technical analysis tests. In this way, the process; will have evolved into a proactive approach with fault prediction. In this way, within the framework of sustainability, benefits will be provided in terms of the environment, economy, and society, and the purpose of supporting stakeholders will be served, especially in terms of time and cost.

The next study will be to detect time-dependent faults by considering the time variable in the data set.

## References

1. Wakiru James, Pintelon Liliane, Analysis of lubrication oil contamination by fuel dilution with application of cluster analysis, Serbia, 2017
2. İbrahim H. Çağlayan, İş Makinalarında Yağ Analizi Ve Önemi, TR, 2003
3. Burak Gökalp, H. İbrahim Saraç, Cenk Çelik, Yağ Analizi Programı İle Aşınmaya Bağlı Hasar Analizi, TR, 2007
4. Wakiru James, Pintelon Liliane, A Decısıon Tree-Based Classıfıcatıon Framework For Used Oıl Analysıs Applyıng Random Forest Feature Selectıon, Serbia, 2018

5. Taşabat E. S., Tayfun Özçay, Salih Sertbaş, Esra Akca, Industry 4.0 Application on Diagnosis Prediction of Construction Machinery: A New Model Approach, TR, 2020
6. Eric R. Scerri, The Periodic Table, USA, 2011
7. Wakiru James, Pintelon Liliane, A Decısıon Tree-Based Classıfıcatıon Framework For Used Oıl Analysıs Applyıng Random Forest Feature Selectıon, Serbia, 2018
8. Tanveer Zafar, Muhammad Shaheen, Sajid Khan, Decision tree classification: Ranking journals using IGIDI, Pakistan, 2019
9. Borusan Makina ve Güç Sistemleri, Web link,https://www.borusancat.com/tr/service/sos-oil-analysis.
10. Akgöbek Ö. and Öztemel E., 2006, Rule Generation Methods Of Inductive Learning Algorithms and Comparisoin of Their Performances, SAÜ Fen Bilimleri Enstitüsü Dergisi, 1, 2006.