# EMPLOYEE TURNOVER PREDICTION

Benan Bardak and Fatih Furkan Has

***Abstract*** **- Nowadays, the business world is continuing to grow every day. Consequently, the number of corporate companies is increasing. All those companies are in a competition in this developing world. One of the most important factors in this competition is the economical state of the company. Success is the most important thing that effects the economical state and the number of qualified members effects the success. Untimely losses of these members effect the companies' success and economy in a negative way. Also, these losses of members can lead to problematic developments of projects and loss of motivation. Because of these reasons, companies do not want to lose their qualified members. This has become the number one priority of companies. To prevent this, human resource department in a company must work closely with the members and increase their motivation and morale. By doing that, companies can analyse the reasons for untimely losses and find a solution. In this data mining Project, our goal is to make analysis to help companies' HR departments. We have basically two targets as a result of working in that dataset: First, to understand and analyse why the employees want to leave the company and second, to predict the employees who are likely to leave the company. We want to share our analysis and predictions to the IT department of the company we are working with, thus; necessary work can be done to prevent employees from leaving and companies can prevent themselves from getting negatively affected economically.**

## I. INTRODUCTION

Databases are increasing and developing every day; as a result of this, the amount of data is increasing. Data's value is recognised more today. In all of the business sectors, data about that workplace is stored in big databases for an amount of time. After readying the data for analysis, data can be analysed. With the help of those analysis, predictions are made about various areas and under these predictions, amenities are provided to the people.

Even if these developments are not noticed by people, their social media activities, usage of internet and phones are being stored in databases. This data is being analysed and predicted by companies to make human life easier. Today, also companies' human resource departments store that kind of data. Companies interpret that data to know more about their employees and to improve the company. With the current conditions, these interpretations are crucial.

In today's developing and changing business life, employees' loyalty to a company is decreasing. To keep the negative effect of employees leaving the company to the projects in a minimum level, human resource departments are working to predict these kinds of situations in advance. To make those works more efficient, data mining algorithms should be used on the employees' data. Employers are used to the leaving employees, but the important thing is keeping the qualified members. It is crucial for project managers and directors to predict these kinds of situations before they happen. Solving problems has a great effect on employees; because of that, human resource departments must work hard on that subject. These kinds of studies are done to find answers to the questions such as: "Why employees quit their job? and "What are the reasons for employees to leave the company?". Some studies show that employers have a great effect on their employees to solve their problems. Employees' reasons to quit their current job may be wanting a better career, industrial accidents, lack of projects or so many projects, work hours, not getting a promotion, financial problems [1]. To predict and solve these kind of issues, human resource departments spend long working hours.

Our study's goal is, by using data obtained by a survey applied on 14999 employees, to find answers to why they left the company and make predictions whether the remaining employees will leave or not and report the reasons to the employers. While making analysis, employees' salaries, average work times, promotions, number of years worked in the company, number of work accidents they had, departments they work in and their happiness will be considered. Data will be prepared to see if those criteria have any effect on employees leaving company. Different analysis and results were gathered using data from similar subjects.

## II. RELATED WORKS

This is an article[2] written about how employees' job attitudes and decisions about leaving the company effects the company's performance and endorsement by looking at the data collected from 911 employees. After this research, it is reached that, employees who are less committed to their companies tended to stay %30 versus %9 after the survey for 18 months. This article was published in 1975 by IBM and it is one of the oldest and most important articles about that subject.

In this article[3], studies have been done about predicting the percentage of leaving because of the leaving employees' negative effects on the company. It is tried to be solved using data mining algorithms. It is aimed that this study's results will help IT departments to solve the problems. The obtained result is, the most three important factors while determining an employee's departure are the mean of the number of token in task report, the standard deviation of working hours, and the standard deviation of working hours of project members in the first month.

In this article[4] , which was published in 2014, the reasons of employee departures were focused on and analysed using data mining algorithms. After these analysis, companies were given

opportunities to solve departure problems. Finding better jobs, haven't been able to move forward in their careers and being unhappy with their salaries are some of the most common reasons for employees to leave a company according to the results.

In another published article[5] , data mining algorithms used for predictions for employees leaving work used by human resource departments are focused on. With the help of these articles, human resource departments are told how they can evaluate employees' futures in the company.

## III. METHODS

### III.A. DATA SET

Researches done on 14999 employees are included in our dataset. There are 10 attributes about 14999 employees. Leaving work percentages of employees working in a company is %24 and satisfaction level of employees working in a company is 0.61. Attributes of our data types are showed in Table 1.

Table 1 – Attribute types of dataset

| | |
|---|---|
| satisfaction_level | float64 |
| last_evaluation | float64 |
| number_project | int64 |
| average_montly_hours | int64 |
| time_spend_company | int64 |
| Work_accident | int64 |
| left | int64 |
| promotion_last_5years | int64 |
| sales | object |
| salary | object |

### III.B. DATA PREPROCESSING

Before making predictions on the data, we must make analysis. Before making analysis on the data, we must modify the data to make it analyzable. For this, we have checked our data to see if there are any missing values. There were not any missing values after our inspections at Table 2.

Table 2 – Missing values of dataset

| | |
|---|---|
| satisfaction_level | False |
| last_evaluation | False |
| number_project | False |
| average_montly_hours | False |
| time_spend_company | False |
| Work_accident | False |
| left | False |
| promotion_last_5years | False |
| sales | False |
| salary | False |

As the result of our analysis, we have updated the "Left" attribute as the last attribute to provide us convenience. Then, we have found the employees' leaving work reasons' percentages. This is done because we are trying to predict their possibility to leave work. As a result, in 14999 employees, 11428 of them remained in the company and 3571 of them left the company. To present it in percentages, the percentage of stayed employees is 0.761917 and the percentage of employees left is 0.238083. By looking at these results, we will try to analyse employees' reasons to leave work.

### III.C. DATA EXPLORATORY ANALYSIS

We have created a correlation matrix to see the power of the relations between the attributes. We have used heatmap structure to visualize correlation matrix (see Figure 1). Expected K-correlation coefficient is between -1 and 1.

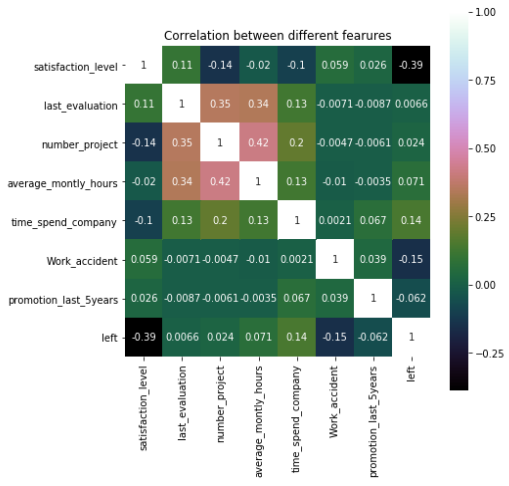


Figure 1 – Heatmap


Figure 2 – Satisfaction level of employees who stayed


Figure 3 – Satisfaction level of employees who left

When we examine the heatmap, we can see that the leaving work percentage "Left" has the most relationships with the satisfaction of the customers. We have seen the correlation coefficient as -0.39 between those two. This means that the more employees are happy, the less they are likely to leave the company. Even if it is not considered as a strong correlation, the couple which has the most correlation coefficient are satisfaction and left attributes.

When the distribution which includes the relationship between the satisfactory levels and the number of employees in the company is examined (see Figure 2), it is seen that employees' who are still in the company, has a satisfaction level on and above average. But by looking at that data, we cannot simply say that if an employee's satisfaction level is high, he/she will not leave the company. When we look at the Figure 3 including left employees and their satisfaction levels, their satisfaction levels are lower and by looking at this result, low satisfaction level employees have a stronger relationship with the employees that have left the company.

While human resource departments are gathering data, they should keep track of how many employees have left work and how many of them are still working in which departments. When we analyse our data, the percentages of employees leaving work are nearly similar between the departments. Sales, Technical and Support departments respectively have the most number of employees. When we analyse these departments, we predict that these departments have the biggest potential for an employee to leave. But, when the ratio of number of employees left and still working to the total number of employees is analyzed, we understand that is not exactly like how we have predicted. For example, it is seen that the number of left employees in human resource departments is more than sales department (see Figure 4). By looking at this, we understand that, number of employees who left the company in a department is not the most important thing; the most important thing is the ratio of employees who left the company to the total number of employees. It is crucial to pay attention to these kinds of analysis while human resource department is making a study to solve problems.

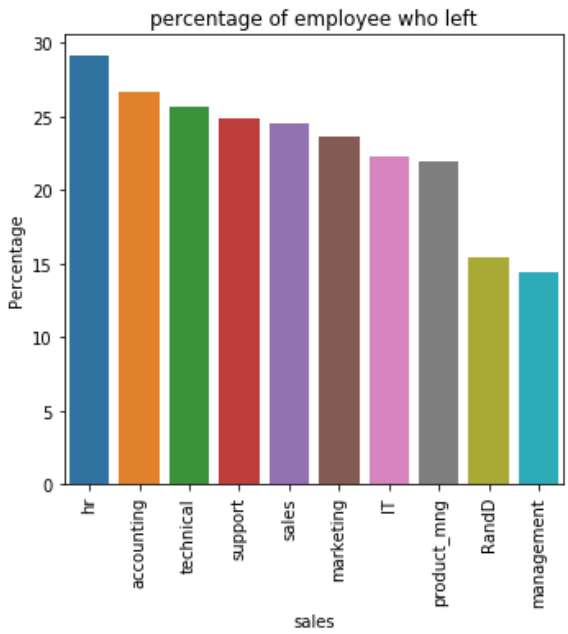


Figure 4 – Percentage of employee who left

For another prediction, we examine the salary-leaving work relationship which is not expected much. It is commonly known that employees with a low salary are more likely to leave work. In our data, salary attribute is grouped as low, medium and high. If we review the salary-leaving work relationship from the table (see Figure 5), we can reach to a result that, employees with a high salary is more likely to stay in the company and employees with a low salary is more likely to leave the company.

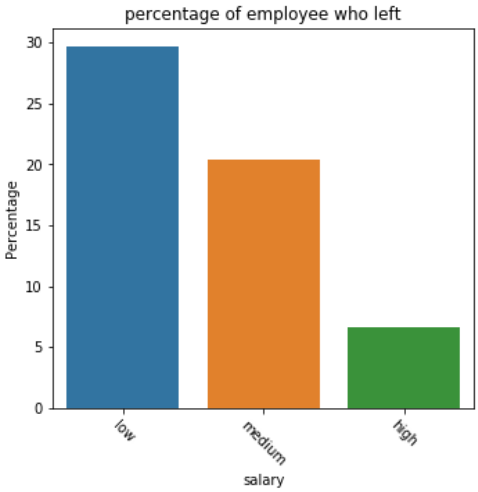


Figure 5 - Salary-Leaving work relationship

The most important reason that we obtained from the correlation matrix is the satisfaction level of an employee is one of the most important factors for the employees who leave work. In the continuation of our work, we wanted to see if the last-evaluation attribute is one of the most important factors. For this, we have sketched a pilot which shows how the satisfaction level and last-evaluation affect the decisions of employees.

When the Figure 6 is examined, it is seen that there are 3 different groups. Firstly, we see a group including employees with satisfaction level between 0.7 and 0.9 and last_evaluation value between 0.8 and 1.0 whom are likely to leave the company. We were not expecting this result, because it can be thought that employees with high satisfaction levels and high last_evaluation values have no problems with the company. But our results show that employees with high satisfaction and last_evaluation go to companies where they can find more opportunities.

In the second group, it is observed that employees with satisfaction level lower than 0.2 and with last_evaluation more than 0.75 are very like to leave the company. By looking at this result, it is important for human resource departments to look into why employees with high last_evalution values have less satisfaction.

The last group on the graphic includes employees with satisfaction levels between 0.35 and 0.45 and last_evaluation values between 0.45 and 0.55. These employees' percentage of leaving work is very high. It is an expected result that unsuccessful employees have less satisfaction and consequently they want to leave the company.

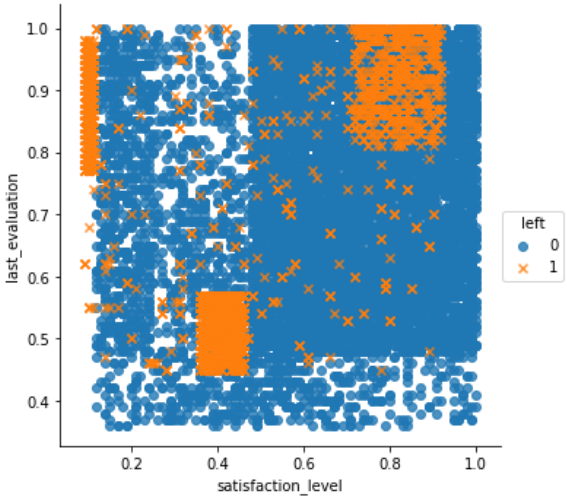


Figure 6 – satisfaction_level-last_evaluation relationship

## IV. RESULTS

In this part, we will get a result by analyzing the data we got from data mining part. We will try to predict who will leave the company by using this result.

## V. DISCUSSION

We will talk about the results, how and where our study can be used and what should be done.

## VI. CONCLUSION

We mill make a summary about the results of the study and what we have done.

## References

[1]Carey & Ogden, 2004; Westcott, 2006

[2]Predicting Turnover of Employees from Measured Job Attitudes ,A L L E N I. K R A U T

[3]5-2017 Who Will Leave the Company? A Large-Scale Industry Study of Developer Turnover by Mining Monthly Work Report Lingfeng BAO ,Zhenchang XING ,Xin XIA,David LO Singapore Management University,

[4]Employee Turnover Analysis with Application of Data Mining Methods
URL http://ijcsit.com/docs/Volume %205/vol5issue01/ijcsit20140501119.pdf

[5]A Data Mining Approach to Employee Turnover Prediction
URL
http://jise.ir/article_10857_380ab2c2c84e1525e1f53647b46d6879.pdf