Correlating Employee Characteristics with Future Tenure in the Company

Metin Akif Zehir, Bilal Emre Durmuş
Computer Engineering Department
Ankara University
Ankara, Turkey
{21290608,22290753}@ogrenci.ankara.edu.tr

Abstract—**This dataset can be used for various HR and workforce-related analyses, including employee retention, salary structure assessments, diversity and inclusion studies, and leave pattern analyses. Researchers, data analysts, and HR professionals can gain valuable insights from this dataset.**

Keywords—employee loyalty, salary structure assesment, human resource

I. INTRODUCTION

In the contemporary landscape of Human Resources (HR) and workforce management, the availability of comprehensive datasets plays a pivotal role in fostering informed decision-making and strategic planning. This article explores a rich dataset that holds immense potential for various HR and workforce-related analyses. The dataset under consideration opens avenues for investigating critical aspects such as employee retention, assessing salary structures, conducting diversity and inclusion studies, and analyzing leave patterns.

As organizations increasingly recognize the significance of data-driven insights in optimizing their human capital, researchers, data analysts, and HR professionals are turning to such datasets for a deeper understanding of workforce dynamics. The wealth of information encapsulated within this dataset enables stakeholders to derive meaningful conclusions and implement targeted strategies for organizational growth and employee satisfaction.

Throughout this article, we will delve into the key dimensions and applications of the dataset, shedding light on its potential to revolutionize HR practices. By harnessing the analytical power embedded in this dataset, professionals across industries can gain valuable perspectives, ultimately contributing to more effective and equitable workforce management.

II. MATERIALS AND METHODS

A. Data Collection

The dataset used for the study was obtained from internal company records. Fundamental demographic and job-related information, such as educational levels, years of joining the company, place of residence, salary levels, ages, genders, ever benched status, and current position experience, were collected to create employee profiles.

B. Data Preprocessing

Initially, we had 4640 instances, but the target values of these instances were not evenly distributed. In order to have a more balanced data set, we used the SMOTE filter, which is a supervised filter that allows generating new artificial data. After applying SMOTE filter, we obtain 6106 instances. Then, we allocated 80% of these for training and 20% for testing.

C. Machine Learning Model

We used the weka program while developing this model. Since our data was the way we wanted, we did not apply any filters to our data. To classify the data, we applied the IBk (kNN) classifier in the Classification tab by providing our test data.

III. RESULTS

According to the results we obtained from the kNN classification which we applied by changing the k's, we observed that the highest ROC value was when k = 1. We observed that when we increased the k values, the ROC value decreased. That's why we set k to 1.
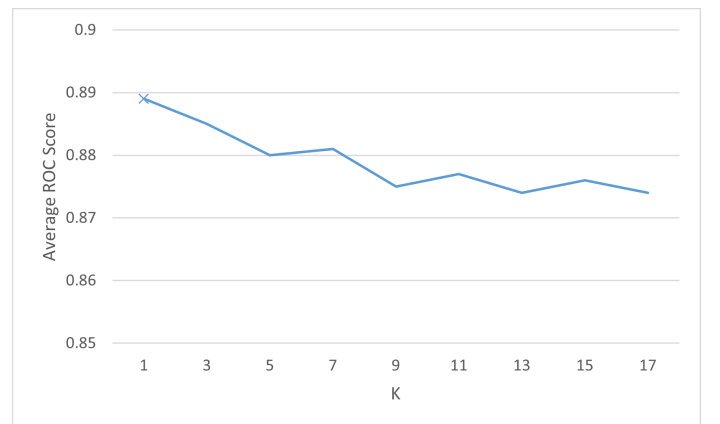

Fig. 1. Finding optimal K

K	TP	FP	FN	TN	Precision	Sensitivity	Specificity	Accuracy
1	0.836	0.165	0.164	0.835	0.835	0.836	0.835	0.836
3	0.815	0.186	0.185	0.814	0.814	0.815	0.814	0.815
5	0.813	0.187	0.187	0.813	0.813	0.813	0.813	0.813
7	0.818	0.183	0.182	0.817	0.817	0.818	0.817	0.818
9	0.813	0.188	0.187	0.812	0.812	0.813	0.812	0.813
11	0.817	0.185	0.183	0.815	0.815	0.817	0.815	0.816
13	0.813	0.189	0.187	0.811	0.811	0.813	0.811	0.812
15	0.812	0.190	0.188	0.810	0.810	0.812	0.810	0.811
17	0.804	0.198	0.196	0.802	0.802	0.804	0.802	0.803

Fig. 2. According to K, other evaluation metrics

IV. CONCLUSION

To sum it up, this dataset is super useful for making HR and work decisions better. It helps in understanding and improving things like keeping employees, deciding salaries, and making workplaces fairer. As companies like to use data for smart decisions, this dataset becomes really important. It gives us good info to help organizations grow and make employees happier.

In a nutshell, this dataset is not just for looking at numbers but can really change how we manage work. It's like a secret weapon for making workplaces better for everyone.

REFERENCES

K. Açıcı and H. Oğul, "Inferring Microarray Relevance By Enrichment Of Chemotherapy Resistance-Based MicroRNA Sets"

Tawfik Hammad, "Employee Data For Classification"