DYNAMIZING HUMAN RESOURCES RECRUITMENT AND EMPLOYEE PERFORMANCE TRACKING WITH BAYESIAN NETWORKS AND IMPROVING SALARY OFFER PROCESSES WITH CLASSIFICATION BASED MACHINE LEARNING MODELS AND CV SCANNING WITH NATURAL LANGUAGE PROCESSING (NLP)



AHMED DEMİREZEN // AHMET CANER SAĞIR // MEHTAP SAATÇI // NECATİ BATUHAN BAYRAKÇI // SAFİYE AYBALA KILIÇ Academic Supervisor: Assist Prof. Rukiye Kaya // Company Supervisor: Fatma Hilal Karatay

INTRODUCTION

HAVELSAN A.Ş., founded in 1982 to meet the technological needs of the Turkish Armed Forces, is a major company headquartered in Ankara, with a strong presence globally. It employs thousands of people, making its human resources operations particularly intense, especially during recruitment phases where pre-evaluation of candidates can be challenging due to high demand. To address this, a decision support system using Bayesian Networks is proposed to pre-evaluate candidates' CVs and assess their suitability for job descriptions. This system also includes post-hire performance tracking for potential promotions and employee performance.

Additionally, a decision support system for salary estimation is proposed to prevent wage inequality, using artificial intelligence to determine appropriate salaries. The project also incorporates a CV scanning system using Natural Language Processing to efficiently review resumes from LinkedIn based on candidates' competencies.



SYSTEM UNDER CONSIDERATION

The company uses the domestic ERP system called "Kovan" in its recruitment operations. Applications to the General Application and Candidate Engineering pools are made through this system, and HR filters the candidates from the 32,000 data pool. This early evaluation is time-consuming as HR manually reviews eligible resumes, forwards them to team leaders, and then selects candidates for interviews. The recruitment process consists of several steps: interview invitations, General Assembly, Human Resources Board, Reference Check and Salary Offer. Without a preliminary examination, all candidates who meet the criteria must be evaluated; This makes the selection process long and complex, taking an average of six months.

PROBLEM DEFINITION

In today's corporate world, HR is vital for performance. However, manual job evaluations and limited application channels hamper diversity and efficiency. The absence of systematic salary determination leads to wage disparities and confusion.

The project proposes a transformative HR approach using Bayesian Networks and AI applications to enhance accuracy, efficiency, and fairness. Bayesian Networks aid in pre-evaluating job applications, assessing candidate suitability, and managing post-recruitment performance. An AI-based salary prediction system ensures fair wage determination. NLP-based resume screening automates the process, ensuring swift and efficient screening of LinkedIn resumes.

RESULTS

A. Results of Preliminary Evaluation of Candidates Before Bayesian Network: Clustering Method

The distribution of 6 clusters for the "System Expert" position is depicted in the figure below, with X points denoting the midpoint of each cluster. 20 Clusters 0 (red), 2 (green), and 5 (yellow) were chosen for the mid-level systems specialist position, 3 15 totaling 429 data points for entry into the Bayesian Network. Other clusters were excluded as they did not align with the intermediate level position criteria. HR can flexibly select clusters for the Bayesian Network based on position level.



B. Results of Bayesian Network Program with Python

Below are the results of a sample candidate that was predicted from the Python program prepared for the Bayesian Network according to the Medium Experienced System Expert Position. The probabilities of all statuses that Exam, Competence, Language, Education, Foreign Language, Foreign Language Exam and Exam scores are shown.





Bayesian Network for Medium Experienced

Probabilistic Display of the Scores for the Medium Experienced System Expert Sample Candidate

 \sim

METHODOLOGY

Looking at general data statistics, there are 137 different positions in total. A total of 1298 competencies are required for these positions. The company has a total of 29,000 application data. The methodology of the project consists of 6 stages:

A. Preliminary Evaluation of Candidates Before Bayesian Network: Clustering Method

Before applying the Bayesian Network, a screening phase using clustering was carried out due to the large amount of data and changing location criteria. This increases the accuracy of the model by ensuring that candidates entering the Bayesian Network meet the required criteria and provides a flexible decision-making system tailored to position-based needs. For this method, the competency scores and years of experience of the relevant candidates are taken as basis. Elbow Method was used to determine the number of clusters.

To simplify calculations, data containing both qualitative and quantitative values are categorized. Categories are based on both the Havelsan Kovan system and commonly accepted standards. The table below provides examples of converting data into categorical data.

C. Bayesian Network and GeNIe Program

Bayesian networks, also known as Belief Networks or Decision Networks, are statistical tools for probabilistic modeling and managing uncertainty. They represent relationships between events using probability theory and help estimate probabilities of outcomes or understand event interdependencies. According to Yang (2019), Bayesian networks are probabilistic graphical models that quantify conditional dependencies among random variables based on Bayes' theorem.

The challenging part of the project was that the Conditional Probability Table sizes were quite large. Therefore, Latent Variables were introduced Beginning to simplify the Bayesian Network structure and Basic Level uthority handle a large number of conditional probabilities. Advanced_Level 0% This approach reduces complexity and facilitates accurate predictions by taking into account unobservable factors. A part of the model obtained eginning with the GeNIe program within the scope of this Basic_Level Authority method is shown in the figure on the right. A total Advanced_Level 0% of 7 latent variables and indicators depending on Exert these nodes were used in the model.

University Institutions Havelsan priority universities Havelsan non-priority universities

Categorical Data for University Institutions

P(B|A)P(A)P(A|B) =P(B)

Bayes Theorem



D. Bayesian Network Program with Python

A Python program with a user-friendly interface for Bayesian Network has been developed to address limitations of the Genie program. The new interface removes constraints like inputting data one by one and enhances usability.

System Expert Position in the Interface

C. Results of Machine Learning for Salary Cluster Prediction

The accuracy rates of machine learning models for predicting salary categories were compared. Due to the large number of salary categories and limited data on extreme values, the accuracy rates of a total of 7 models are in the range of approximately 0.50 to 0.60. These accuracy rates seem sufficient for the current problem. With increasing data, model performance may improve upon rerun. Considering detailed analysis, the KNN model was preferred for the interface due to its balanced performance across training and testing datasets, shorter runtime, fewer parameters to control, and lower memory usage. The figure on the right shows the interface created for the user.



Display of the Predicted Category in the Interface

D. Results of Natural Language Processing (NLP) for LinkedIn Integration

| 0 | CV NLP Analysis | | | - | | \times | | | |
|------|------------------|---------------------|-----------------|---|--|----------|--|--|--|
| elec | ct Position: | | | | | | | | |
| /eri | tabanı Mühendisi | | | | | ~ | | | |
| | | | Select Folder | | | | | | |
| | Run Analysis | | | | | | | | |
| _ | | | 100% | | | | | | |
| | PDF Name | NLP Score | | | | <u></u> | | | |
| 1 | M.pdf | 0.1891891891891892 | | | | | | | |
| 2 | G.pdf | 0.16216216216216217 | | | | | | | |
| 3 | O.pdf | 0.13513513513513514 | | | | | | | |
| \$ | Apdf | 0.10810810810810811 | | | | | | | |
| 5 | B.pdf | 0.10810810810810811 | | | | | | | |
| 5 | P.pdf | 0.08108108108108109 | | | | | | | |
| 7 | Lpdf | 0.05405405405405406 | | | | | | | |
| 8 | C.pdf | 0.0 | | | | | | | |
| 9 | D.pdf | 0.0 | | | | | | | |
| 10 | E.pdf | 0.0 | | | | | | | |
| 11 | F.pdf | 0.0 | | | | | | | |
| | | | Show Graph | | | | | | |
| | | | Show Skill List | | | | | | |
| | | | Save to Excel | | | | | | |

Developed Python interface for resume screening system using Natural Language Processing. The user selects the position and uploads a file in PDF format containing resumes. Any number of PDFs can be uploaded to the system. After running the analysis, the results are displayed, including skills extracted from the resumes and similarity rates. Users can save this information as a text file and edit competency skills if necessary. Jaccard Similarity results are given for Similarity Score. The figure on the left shows the results obtained within the interface.

Similarity Score Results Screen in the Interface

CONTRIBUTIONS PROVIDED TO THE COMPANY

Within the systems, the recruitment rate will increase in the long term through the Bayes Network system. Thanks to the systems, the turnover rate will decrease and HR processes will become shorter and more effective. With the salary estimation system, the number of revisions will be reduced and the acceptance rate of offers will increase. In the short term, two evaluations were made to evaluate the outcomes of the programs.

-The project compared Bayesian Network's resul rankings provided by Have specialists for three position

| d the | Position | Candidates | Differently Ranked | Match Rate |
|----------|-------------------------|------------|--------------------|------------|
| ts with | System Expert | 30 | 8 | 73.30% |
| elsan HR | Software Engineer | 36 | 8 | 77.70% |
| ons. | System Integration Eng. | 41 | 12 | 70.70% |

-The project achieved substantial time and cost savings:

The new candidate selection system of Bayesian Network significantly boosts efficiency. Previously, manual filtering

Ζ

 $\mathbf{\alpha}$

ט Z

ш

INDUSTRIAL

 \bigcirc

Ζ

ш

Σ

2

Predict Data

HAVELSAN



ENGINEERING

A user-friendly Python program for Bayesian Network has been developed to overcome limitations of the Genie program. It allows for importing categorical data, creating nodes, network connections, establishing and saving/loading networks. Users can modify network shape, run calculations, upload evidence collectively via excel, and predict Bayesian Network results. Predicted data is output as excel, facilitating candidate percentage viewing. The figure on the left shows the interface screen.

Latent Variable for Competence Score and its Indicators

E. Machine Learning for Salary Cluster Prediction

The dataset used for machine learning consists of employee information with a total of 1820 entries with 31 columns. 30 of these are independent variables and 1 is a dependent variable. Some features were directly obtained from Havelsan, while others were engineered by us. After examining correlations with the dependent variable "Salary Given Category," 11 independent variables were selected. These variables and their correlations are detailed in the figure below. "Salary Given Category" represents salary scales provided, categorized by Havelsan, with 7 undisclosed categories indicating increasing salary ranges.

required 9 hours to select 10 candidates. With the new system, this time is reduced to just 45 minutes, saving 91.67% of

| System (Recruitments per year) | Expected Annual Income (TL) | | |
|--------------------------------|-----------------------------|--|--|
| Bayesian System | 412,384.5 - 549,846 | | |
| Machine Learning System | 68,400 - 102,600 | | |
| NLP System | 8,750 - 17,500 | | |
| Total | 489,534 - 669,946 | | |

CONCLUSION

The project aimed to streamline HR candidate evaluation using Bayesian Networks, focusing on experience, education, and language skills. Clustering and categorical data transformation simplified the process. Python coding and an interface handled large data volumes efficiently. Additionally, a cluster-based salary forecasting system and a resume scanning system using Natural Language Processing were developed. These solutions reduced Havelsan's workload, saved time and money, and provided effective HR management.

F. Machine Learning for Salary Cluster Prediction

Automating resume screening with NLP streamlines the process and reduces biases. Key steps include extracting relevant information, preprocessing resumes, and measuring competency match using Jaccard Similarity. Cosine Similarity, although explored, was less suitable due to its focus on word frequency. Jaccard Similarity proved effective, indicating stronger candidate alignment with job requirements through higher scores.

References:

the time.

[1] Yang, X.-S. (2019). Mathematical foundations. Introduction to Algorithms for Data Mining and Machine Learning, 19-43. doi:10.1016/b978-0-12-817216-2.00009-0