# Analysis of work behavior among employees at the University of Hafr Al-Batin

**fadiyah alanazi , wegdan alharthi , elham khatim , ibtesam falih ,ghazlan alanazi**
**Depatment of Computer Science and engieering**
**Supervisor: Dr. Abdulrahman Alzahrani**

**كلية علوم وهندسة الحاسب**
College of Computer Science and Engineering
University of Hafr Al Batin  جامعة حفر الباطن

**جامعة حفر الباطن**
University of Hafr Al Batin

## Abstract

This project, titled "Analysis of the Work Behavior of Hafar Al-Batin University Employees," aims to explore and analyze the intricate relationship between employee work behavior and various demographic factors such as gender, educational qualifications, rank, and job title. Employing advanced data analysis techniques and machine learning algorithms, the study seeks to understand how these factors influence employee discipline during working hours.

The research utilizes a comprehensive dataset from Hafar Al-Batin University, incorporating various employee demographics and work behavior records. Machine learning models, particularly classification algorithms, are applied to identify patterns and correlations within this data. The study also integrates a dashboard for visualizing and interpreting the results, facilitating a clearer understanding of the impacts these demographic factors have on employee discipline.

The outcomes of this research are expected to offer valuable insights for the university's administrative and HR departments, aiding in the development of more effective policies and strategies for employee management. Furthermore, the findings could contribute to the broader academic understanding of work behavior in educational institutions, providing a template for similar studies in other contexts.

This project represents a significant step towards integrating data-driven approaches in the analysis of organizational behavior, highlighting the potential of machine learning and data analytics in human resource management and organizational studies.

## Introduction

General introduction:
The study of employee behavior within university settings is crucial for optimizing administrative and academic operations, fostering a productive work environment, and enhancing employee satisfaction. This research focuses on analyzing several factors that influence employee behavior at the University of Hafr Al-Batin, utilizing a data-driven approach to uncover underlying patterns and determinants.

Importance of Study :
Understanding the dynamics of employee behavior is vital for higher education institutions that strive to maintain an efficient and harmonious workplace. This study is particularly significant as it addresses the need for a comprehensive analysis of employee behavior in the context of a Saudi Arabian university, which is often underrepresented in global research. The findings can provide valuable insights into the management strategies that can be employed to enhance employee performance and job satisfaction.

Problem Statement:
Despite the critical role of employee behavior in the success of academic institutions, there is a gap in systematic and data-driven analysis concerning how various demographic and job-related factors influence such behaviors in the context of universities in the Middle East, particularly Saudi Arabia. This gap hinders the ability of university administrators to implement effective strategies tailored to their unique organizational culture and workforce.

Aims and Objectives :
The primary aim of this project is to analyze and understand the factors influencing employee behavior at the University of Hafr Al-Batin. The specific objectives include:
To identify key demographic and job-related factors affecting employee behavior.
To utilize statistical and machine learning techniques to analyze these factors.
To provide recommendations based on the analysis that can help improve employee management and policymaking.
Identifying key factors that influence employee behavior.
Understanding the relationship between these factors and overall employee performance and satisfaction.
Providing data-driven recommendations to the university's management for policy and strategy development.

Scope and Limitations of study:
This study will focus on analyzing available data from the University of Hafr Al-Batin's employee records, considering variables such as gender, age, educational background, job title, and department. While aiming for a comprehensive analysis, the project acknowledges limitations in data availability and the scope of factors considered. The findings are intended to offer insights and recommendations specific to the University of Hafr Al-Batin but may also be relevant to similar institutions seeking to enhance their understanding of employee behavior.

## Background

At the heart of organizational effectiveness in educational settings is the understanding and optimization of employee behavior. The University of Hafr Al-Batin faces unique challenges and opportunities typical of higher education institutions in Saudi Arabia, where cultural and organizational dynamics play a significant role in day-to-day operations.
Recognizing the lack of comprehensive data-driven studies in this regional context, our research aims to fill this gap by analyzing how various demographic and job-related factors influence employee behavior at the university.
Utilizing advanced data analysis and machine learning techniques, this project seeks to uncover patterns and determinants of employee behavior that are often overlooked in traditional studies. By focusing on specific factors such as gender, age, educational background, and job title, the study provides insights that are crucial for the development of effective management strategies and policies tailored to the unique cultural and organizational context of Middle Eastern academic institutions.
The ultimate goal is to leverage these insights to foster a more productive work environment and enhance employee satisfaction, thereby improving overall institutional efficiency and effectiveness. This research not only benefits the University of Hafr Al-Batin but also contributes to the broader academic discourse by providing a template for similar analyses in other regions and contexts.

## Methodology

Environment Setup:
The analysis was conducted using several key tools:
Programming Environment: Python 3.8, an excellent choice for data science due to its robust libraries and community support.
Development Tools: google colab, Jupyter Notebook, which provides an interactive coding environment ideal for data exploration and visualization.
Python library: Pandas for data manipulation and analysis , and matplotlib for creating static, animated and interactive visualization and seaborn is a statical data visualization based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
Google Colab: is an improved cloud version of Jupyter Notebook dedicated to writing and running code and Notebook documents. It helps you write and execute code written in Python through your browser without the need to install any editor or program.
ML/Deep Learning Model :
Through the data obtained, we created Sample Machine Learning, which is a simple sample model that explains how to use Sample Machine Learning using scikit-learn, through which we used what is known as Random Forest. Since our data does not strongly support machine learning, this model was the best fit after conducting several experiments using other models. scikit-learn it is a library in the Python language that provides many supervised and unsupervised learning algorithms. It is built on some technologies that you may already be familiar with, such as NumPy, pandas, and Matplotlib. It is characterized by being a consistent interface for machine learning models. Random Forest is a powerful and versatile supervised machine learning algorithm that grows and combines multiple decision trees to create a "forest" that works on the principle of working in parallel rather than sequentially, which in turn helps improve results and connect to higher accuracy.
Dataset Evaluation :
In the initial phase of our project, we submitted a request to the university to access the employment data for use in our project. This request was approved, and we were granted access to the data, which consisted of two files. The first file contained employee information, and the second comprised their fingerprint data. We merged these two files using a unique identifier, resulting in a consolidated dataset. This comprehensive data file now contains information organized into 10 columns and encompasses a total of 73,035 entries.
The dataset includes the following column names, which facilitate detailed analyses and categorization of the employee data:
ID: A unique identifier assigned to each employee.
Date Time: The timestamp associated with each recorded entry.
Type: Classification of entries, detailing the nature of the data recorded.
M: A code representing specific metrics or categories relevant to the employee data.
Sex: The gender of the employee.
Formation: Details pertaining to the employee's formative background or grouping within the organization.
Rank: The official rank of the employee within the organization.
Class: Classification based on employment or administrative criteria.
Job Title: The official title of the employee's position.
Qualification: The academic or professional qualifications held by the employee.
We will proceed to analyze this dataset and present the findings. The results will be visualized through an interactive dashboard, which will allow for dynamic interaction with the data. This visualization will enable stakeholders to easily comprehend the trends and patterns within the employee data, facilitating informed decision-making and strategic planning.
Training for ML/ Deep Learning Model :
The data was divided into features and objectives, training, and testing, then we used both Random Forest and K-Nearest Neighbors (KNN) models. We found that Random Forest was 0.73% better compared to K-Nearest Neighbors (KNN), which was 0.73% less efficient. 0.69.
Evaluation Metrics of Prediction Model:
Since the project relies on classification models, it needed evaluation standards that were appropriate and compatible with it, the most prominent of which were used as follows:
Accuracy: Accuracy measures the proportion of true results (both true positives and true negatives) among the total number of cases examined.
Precision: Precision is the ratio of correctly predicted positive observations to the total predicted positives. It is a measure of a classifier's exactness.
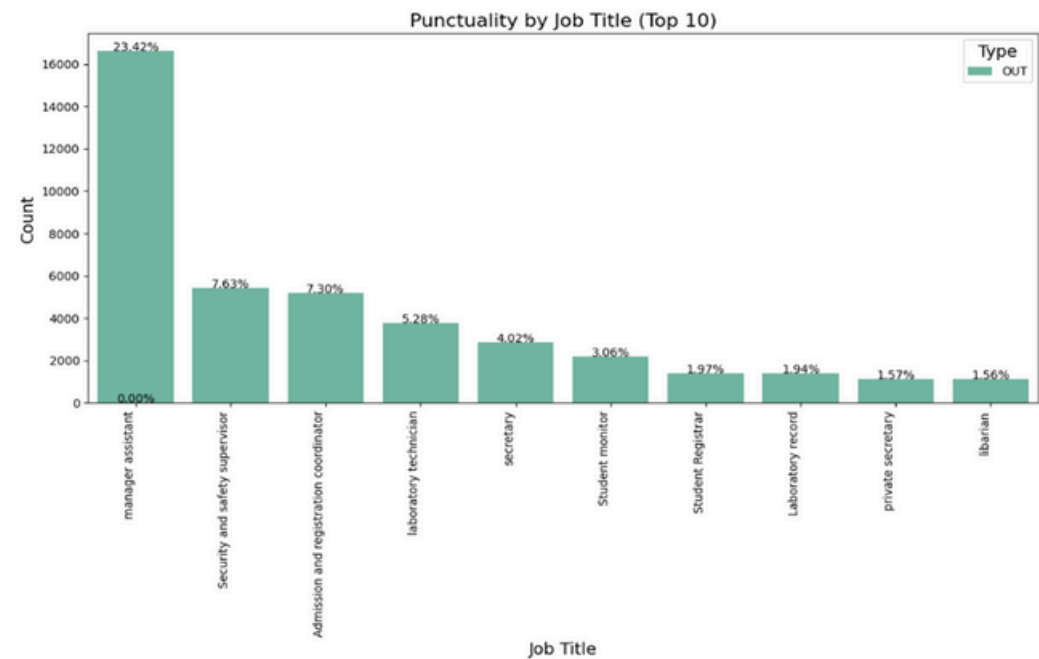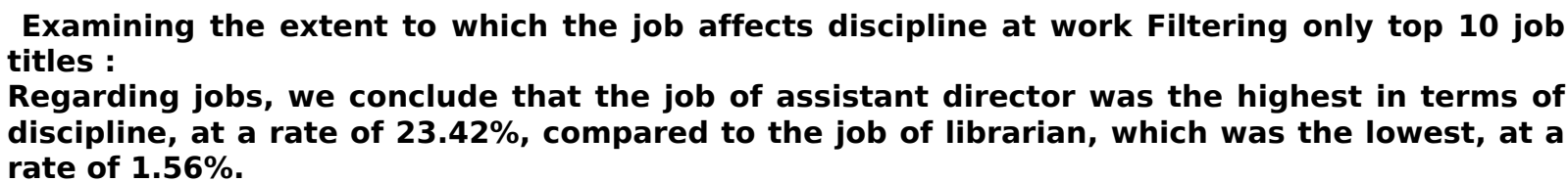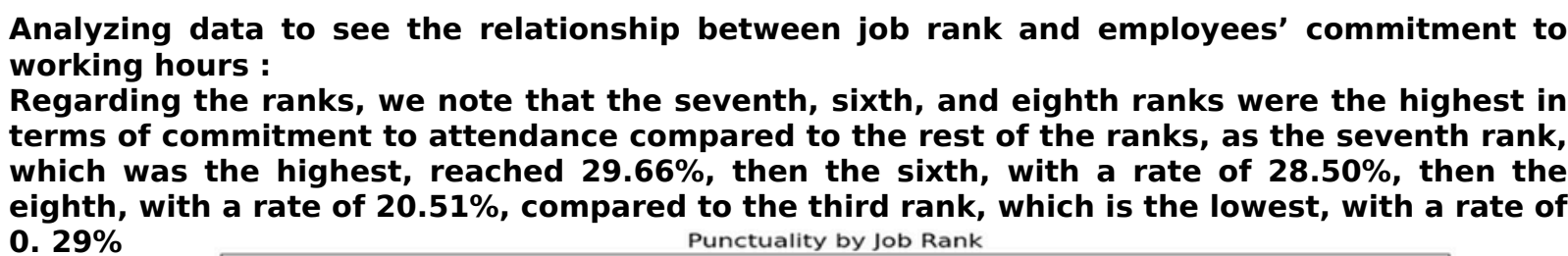Recall: Recall is the ratio of correctly predicted positive observations to all observations in actual class - yes. It is a measure of a classifier's completeness.
F1-score: The F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. It is a balance between Precision and Recall

## Results and Evaluation

The application of the machine learning model was made using a random forest identifier in order to create an evlution such as Accuracy - Precision - Recall - F1 - score . Therefore, after completing the implementation, you conclude that the model achieved 70% accuracy, and this means that the model can create a classification of data with a percentage 70% correct and 30% incorrect. Therefore, this value is not the best possible, and this is due to the data we have, as it is not very compatible with machine learning. Therefore, we went to create a visualization and display of the percentages, which in turn will be used to make predictions in the future and make decisions based on them.
Results from the visualization process:
A data visualization was created to reach conclusions and present them more clearly so that they can be easily understood which in turn will have a major role in the decisions that will be made in the future. Analyzing the regularity of attendance and departure based on academic qualification: Through the figure related to qualifications, we conclude that the bachelor's qualification was the highest at a rate of 49.11% compared to the doctorate, which was the lowest at a rate of 0.17% in terms of punctuality.



Comparison of attendance and departure patterns between males and females:
We note that females were more committed to attendance compared to males, with a total of 52.79% of females and 47.21% of males.



Analyzing data to see the relationship between job rank and employees' commitment to working hours :
Regarding the ranks, we note that the seventh, sixth, and eighth ranks were the highest in terms of commitment to attendance compared to the rest of the ranks, as the seventh rank, which was the highest, reached 29.66%, then the sixth, with a rate of 28.50%, then the eighth, with a rate of 20.51%, compared to the third rank, which is the lowest, with a rate of 0. 29%



Examining the extent to which the job affects discipline at work Filtering only top 10 job titles :
Regarding jobs, we conclude that the job of assistant director was the highest in terms of discipline, at a rate of 23.42%, compared to the job of librarian, which was the lowest, at a rate of 1.56%.



## Conclusions

In conclusion, this project was trying to explore the regularity of employees' attendance and departure based on a set of data: gender, rank, qualification, and job title, which in turn is of great importance in helping institutions and bodies discover the most committed employees and the most non-committal ones, which will help later. To appoint the most committed employees. From a technical standpoint, this was done by analyzing and visualizing the data and developing a different set of visualizations while clarifying the ratios on each form to find the most committed people, in addition to applying machine learning, which is Random Forest, which succeeded in reaching the results of Accuracy: 0.73, Precision: 0.74, Recall: 0.73, F1 Score: 0.72. Therefore, this technology provided a lot of information on the subject that may meet the needs of organizations and bodies when making decisions in the future.
Based on the findings, the following recommendations are made for the university's management:
Enhance Diversity Programs: Implement and expand diversity and inclusion training.
Review Work Policies: Update work policies for flexibility and fairness.
Expand Training Opportunities: Offer more professional development to support career growth.
Regular Evaluations: Conduct periodic evaluations to assess policy effectiveness and adapt as needed.

## Acknowledgement

## References

·Bass, B. M., & Avolio, B. J. (1994). Improving organizational effectiveness through transformational leadership. Thousand Oaks, CA: Sage.

·García-Morales, V. J., Jiménez-Barrionuevo, M. M., & Gutiérrez-Gutiérrez, L. (2012). Transformational leadership influence on organizational performance through organizational learning and innovation. Journal of Business Research, 65(7), 1040-1050.

·Herzberg, F., Mausner, B., & Snyderman, B. B. (1959). The motivation to work (2nd ed.). New York, NY: John Wiley & Sons.

·Johnson, S., & Smith, P. (2013). Gender differences in workplace behavior and conflict resolution. Journal of Business Studies Quarterly, 5(1), 12-22.

·Kotter, J. P., & Heskett, J. L. (1992). Corporate culture and performance. New York: Free Press.

·Leith wood, K., & Jantzi, D. (2000). The effects of transformational leadership on organizational conditions and student engagement with school. Journal of Educational Administration, 38(2), 112-129.

·Maslow, A. H. (1943). A theory of human motivation. Psychological Review, 50(4), 370-396.

·Nguyen, T., Taylor, J., & Bradley, S. (2016). Job autonomy and job satisfaction: new evidence. Education Economics, 24(3), 282-307.

·Smith, M., & Robertson, J. (2015). Demographic influences on employee satisfaction in higher education. Teaching in Higher Education, 20(5), 513-527.

·Turner, B. A., & Müller, R. (2005). The influence of project managers on