**Capstone-2: HR Management - Knowledge Graphs**

**Mira Hussien – Mirna Kordab – Sara Wehbe**

**1. Choosing the Dataset**

**Datasets Referred:** The two datasets used in the project were from the book Predictive Analytics in Human Resource Management: A Hands-on Approach by Shivinder Nijjer and Sahil Raj. Both the datasets are related to employee information data and are in CSV file format with 584 rows of data.

The datasets can be accessed here:

1. <https://drive.google.com/file/d/1hkKNFty5mdW1fxv0m-B18mH5tJmYD2L-/view?usp=sharing>
2. <https://drive.google.com/file/d/1TZ_6PJuyArPQl84RA3VUuiwLCKjrgO3c/view?usp=sharing>

This selection of datasets is important for our project since they add richness and details of information on employees that would be ideal in analyzing performance, job satisfaction, and turnover.

**Dataset Overview:**

1**. Predictive Model of Selection Dataset:** This dataset will analyze a number of factors that predict the performance and satisfaction of employees. It encompasses a number of psychological characteristics and information on employee backgrounds, including:

* **Marriage Status:** Whether married or single.
* **Gender:** The gender of the employee.
* **Psychological Traits:** Scores on integrity, efficacy, humility, esteem, resilience, extroversion, agreeableness, conscientiousness, emotional stability, and openness.
* **Performance Indicators:** Ratings for employee performance that reflect academic grades and overall job satisfaction.

2. **Predictive Model Turnover Dataset:** This dataset speaks to the likelihood of turnover, which means leaving the organization. The dataset captures similar demographic variables and metrics such as:

* **Person-Job Fit (PJ\_Fit):** It indicates the degree to which a person fits his or her job role.
* **Performance Rank (Perf\_Rank):** The ranking on performance for an employee within the organization.
* **Job Satisfaction (JobSat):** It gives the score on attitude towards satisfaction with their job.
* **Quality of Work Life (QoWL):** The quality of work life an employee leads within the organization.
* **Intention to Leave (IntentionCode):** It provides the code for the likelihood of leaving the job.

**Relevance to the problem of HR Management:** These datasets are very relevant for our project in HR management because they will provide the means for investigating and visualizing the relationships that exist between employee characteristics and key outcomes such as job performance and turnover. Therefore, it would be possible by drawing inference from the analysis to come up with appropriate insights that can help HR professionals make informed decisions on how to retain employees and improve performance.

**Entities and Attributes:**

• **Entities:** In each of the datasets, we have 584 rows, with each one representing one employee's set of data. This suggests that we are beginning to have enough data from which some meaningful insight could be derived, which is an important requirement when drawing a meaningful conclusion.

**1. Predictive Model of Selection Dataset - Codebook**

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| **Marriage** | Indicates marital status (1 = Married, 2 = Single) |
| **Gender** | Gender of the employee (1 = Male, 2 = Female) |
| **Integrity** | Score reflecting the integrity of the employee (on a scale from 1 to 5) |
| **Efficacy** | Score measuring the self-efficacy of the employee (on a scale from 1 to 5) |
| **Humility** | Score for humility (on a scale from 1 to 5) |
| **Esteem** | Score measuring self-esteem (on a scale from 1 to 5) |
| **Resilience** | Score measuring resilience (on a scale from 1 to 5) |
| **Extrovert** | Score measuring extroversion (on a scale from 1 to 5) |
| **Agreeable** | Score for agreeableness (on a scale from 1 to 5) |
| **Conscience** | Score for conscientiousness (on a scale from 1 to 5) |
| **Emotions** | Score reflecting emotional stability (on a scale from 1 to 5) |
| **Openness** | Score measuring openness to new experiences (on a scale from 1 to 5) |
| **Performance** | Rating of employee performance (on a scale from 1 to 5) |
| **Grades** | Academic or performance grades of the employee |
| **Job Satisfaction** | Overall job satisfaction score (on a scale from 1 to 5) |

**2. Predictive Model of Turnover Dataset - Codebook**

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| **Marriage** | Indicates marital status (1 = Married, 2 = Single) |
| **Gender** | Gender of the employee (1 = Male, 2 = Female) |
| **PJ\_Fit** | Score indicating how well the employee fits their job (on a scale from 1 to 5) |
| **Perf\_Rank** | Ranking based on employee performance (on a scale from 1 to 5) |
| **PO\_Fit** | Score indicating how well the employee fits with the organization (on a scale from 1 to 5) |
| **JobSat** | Overall job satisfaction score (on a scale from 1 to 5) |
| **QoWL** | Score reflecting quality of work life (on a scale from 1 to 5) |
| **IntentionCode** | Code indicating the likelihood of the employee leaving the organization (1 = Low, 2 = High) |

**2. Describing Future Analyses to Be Done Later**

**Stakeholders Interested in the Analyses:** Several stakeholders within an organization will find our analyses of great use:

* **HR Managers:** Especially interested in learning about what contributes to employee satisfaction and retention, since such information might be useful for the formulation of policies to help improve such problems.
* **Organizational Leaders:** Will benefit from findings that would inform strategic decisions with respect to workforce management and development.
* **Data Analysts:** will ensure that the insights coming from the data are validated and the methodologies applied are sound and applicable.

**Intelligent Analyses Planned:**

1. **Graph Analytics through Knowledge Graphs:** We intend to create knowledge graphs to represent graphically how various employee attributes inter-relate. For instance, we can depict how a number of psychological traits relate both to job satisfaction and to turnover intentions, allowing us a fuller picture of employee dynamics.

2. **Predictive Regression Models:** We would implement regression analysis to predict certain attributes like job satisfaction and intention of turnover based on other attributes in the datasets. These will help us identify the key determinants for the same.

3**. Classification Models:** We look forward to testing some classification techniques that will enable us to classify employees as those likely to leave or those performing well. This will assist HR managers in being ahead in saving potential turnover issues.

4**. Data Exploration for Actionable Intelligence:** EDA will be done to look out for trends, patterns, and insight that could help the human resource strategies. This may include checking the correlations across various attributes to find out the underlying factors contributing to employee performance and satisfaction.

**Technologies to Be Used:** Implementation of the analyses will be with the help of the following technologies:

* **Programming Languages:** Python and R are our mainstays for handling, analyzing, and visualizing data.
* **Database Management:** We will be using SQL in managing and querying our data sets with efficiency.
* **Machine Learning Libraries:** Scikit-learn for regression and classification, TensorFlow for more complicated predictive model building.
* **Graph Databases:** Neo4j will be used to build knowledge graphs, graphically depicting the relationships among these data.

**Convincing Stakeholders of Analysis Value:** Relevant case studies will be introduced to help stakeholders understand the value of our proposed analyses, where companies have witnessed better employee retention and performance outcomes using methods similar to what we propose. Emphasize also on potential cost savings from reduced turnover and improved employee engagement.

**Cost-Benefit Scenarios:** Costs to be considered against benefits when stakeholders contemplate this analysis may include the following:

* Investment in technology and tools for such analysis
* Training for HR people to interpret and act upon insights from such analysis
* Maintenance of the analytics systems.

These can easily be outweighed, however, by benefits accruing from reduced turnover rates, increased productivity of human resources, and job satisfaction that contributes to positive health of the organizational environment.

**Challenges to Deployment:** Some challenges that can be considered an obstacle toward deploying our proposed solutions include the following:

* **Cultural Resistance to Change:** Employees and management are likely to feel shocked by these new data-driven approaches to managing HR. There is an urgent requirement to develop a culture that welcomes insights driven by data.
* **Operational Challenges:** The integration of new analytics processes into the current workflow disrupts routine business operations.
* **Infrastructure Challenges:** Very often, organizations will need to invest in technology upgrades to handle advanced analytics.
* **Readiness Challenges**: Organizational readiness, especially the skills of the HR team about advanced analytics, is very important and has to be assessed.
* **Privacy and Compliance Challenges:** In implementing HR analytics, ethical and legal processing of employee information is paramount.

Using these datasets and knowledge graphs will help our project improve human resource management practices, which, in turn, will lead to improved satisfaction and better retention of employees.