**Capstone 3: Employee Satisfaction and Turnover Analysis Using Predictive Modeling and Graph Analytics**

**1. Introduction**

In this project, we aim to leverage employee data to analyze key aspects such as **job satisfaction** and **employee turnover**. By utilizing two datasets provided in the book "Predictive Analytics in Human Resource Management" by Shivinder Nijjer and Sahil Raj, we apply machine learning and graph analytics to gain insights into the factors that contribute to employee satisfaction and retention.

**2. Datasets Used**

**Datasets Overview**:  
We used two datasets related to employee information:

1. **Predictive Model of Selection Dataset**: Contains attributes related to employee characteristics such as gender, marital status, psychological traits (integrity, efficacy, etc.), performance indicators, and job satisfaction scores.
2. **Predictive Model Turnover Dataset**: Contains variables such as job satisfaction, person-job fit, performance rank, quality of work life, and intention to leave the company.

Both datasets contain 584 rows, with each row representing a unique employee's data.

**3. Research Questions**

The research focuses on answering the following questions:

* **Job Satisfaction**: What factors contribute to employee job satisfaction and performance?
* **Employee Turnover**: Which factors predict the likelihood of an employee leaving the company?

**4. Intelligent Analyses Implemented**

**a. Regression Analysis (Predicting Job Satisfaction)**

**Regression Model**:  
We implemented a **Linear Regression** model to predict **Job Satisfaction** based on several employee characteristics. This model helps us understand which factors are most important in determining employee satisfaction.

* **Regression Results**:
  + **MSE (Mean Squared Error)** = 0.7571
  + **R² (R-squared)** = 0.549

These results suggest that approximately **55%** of the variance in job satisfaction can be explained by the model. While not perfect, the model provides useful insights into factors affecting job satisfaction.

**b. Classification Analysis (Predicting Employee Turnover)**

**Classification Model**:  
We used a **Random Forest Classifier** to predict whether an employee is likely to leave the company (intention to leave) based on attributes such as job satisfaction, person-job fit, and performance rank.

* **Classification Results**:
  + **Accuracy** = 90.3%
  + **F1 Score** = 0.903

The high **accuracy** and **F1 score** indicate that the model is effective in identifying employees likely to leave, which can help HR professionals target retention strategies.

**c. Graph Analytics (Knowledge Graph)**

**Knowledge Graph**:  
We used **NetworkX** to create a **Knowledge Graph** that visualizes the relationships between employee attributes. The graph is based on the correlation matrix of the dataset, with edges representing significant correlations (greater than 0.5) between features.

* **Graph Insights**:  
  The graph reveals that certain psychological traits, such as **Resilience** and **Efficacy**, are closely connected to both **job satisfaction** and **turnover intentions**. This graph provides HR managers with a visual understanding of how employee attributes interrelate, potentially aiding in decision-making processes related to employee development and retention.

**d. Unsupervised Learning (Clustering Analysis)**

**Clustering Analysis (K-Means)**:  
We used **K-Means clustering** to identify natural groupings within the dataset based on attributes like performance and job satisfaction. This allows us to segment employees into different clusters, which could be helpful for personalized HR strategies.

* **Clustering Results**:
  + The **K-Means model** identified 3 distinct clusters of employees. These clusters represent groups of employees with different levels of job satisfaction and performance.

**5. Results Evaluation**

**Quantitative Evaluation:**

* **Regression Analysis**:
  + **MSE** (0.757) and **R²** (0.549) suggest a moderate model that could be improved with additional features or more complex models.
* **Classification Analysis**:
  + **Accuracy** (90.3%) and **F1 Score** (0.903) indicate that the classification model is robust and well-suited to predict turnover.
* **Clustering Analysis**:
  + Clustering coefficients were calculated, revealing which features contribute most to clustering behavior (e.g., **Efficacy**, **Resilience**).

**Qualitative Evaluation:**

* **HR Managers**: These models can assist HR managers in understanding the key factors affecting **employee satisfaction** and **retention**. The **classification model** can help HR professionals identify employees at high risk of leaving, while the **regression model** offers insights into factors that improve job satisfaction.
* **Organizational Leaders**: These analyses can inform strategic decisions, such as **personalized employee engagement** strategies based on clusters of employees with different satisfaction levels.
* **Data Analysts**: The **knowledge graph** and clustering analysis provide a deeper understanding of employee dynamics, aiding in the development of more accurate predictive models and better human resource strategies.

**6. Visuals and Interpretation**

**a. Knowledge Graph**

The **Knowledge Graph** visualizes the relationships between employee features such as **Efficacy**, **Resilience**, **Job Satisfaction**, and **Turnover Intentions**. The graph helps identify which features are strongly correlated, offering a visual representation of the factors influencing employee dynamics.

**b. Clustering Coefficients by Features (Bar Chart)**

The **Clustering Coefficients** bar chart illustrates which features are most influential in clustering employees. Features such as **Efficacy**, **Resilience**, and **Openness** have high coefficients, indicating that they are key in differentiating employee clusters.

**c. Heatmap of Clustering Coefficients**

The **heatmap** of clustering coefficients further visualizes how each feature contributes to the clustering process. Features with higher coefficients are shown in more intense colors.

**7. Conclusion and Future Work**

In conclusion, the **Predictive Model of Selection Dataset** and **Predictive Model Turnover Dataset** provided valuable insights into the factors affecting **job satisfaction** and **employee turnover**. The regression and classification models, along with graph and clustering analyses, offer a comprehensive view of employee dynamics, helping HR professionals and organizational leaders make informed decisions.

**Next Steps**:

* **Model Improvements**: We plan to enhance the models by exploring additional features and using more advanced techniques such as **XGBoost** or **Neural Networks**.
* **More Complex Graph Analytics**: Further analysis of the **knowledge graph** can include **link prediction** and **community detection**, which could yield deeper insights into employee behavior patterns.
* **Integration with HR Systems**: Integrating these models into HR management systems would allow for real-time predictions and targeted employee retention strategies.

**8. Google Colab Link**

You can view and replicate the experiments by accessing the following **Google Colab link**: **Colab Notebook Link** *(Replace with your actual Colab link)*

**9. References**

* Nijjer, S., & Raj, S. (2021). *Predictive Analytics in Human Resource Management: A Hands-On Approach*. Routledge.