UGANDA CHRISTIAN UNIVERSITY
FACULTY OF ENGINEERING DESIGN AND TECHNOLOGY

DEPARTMENT OF COMPUTING AND TECHNOLOGY

Course: MASTER OF SCIENCE IN DATA SCIENCE AND
ANALYTICS
Course unit: AI_ML
PROJECT PROPASAL

GROUP MEMBERS

NAME	REG NUMBER	ACCESS NUMBER
PAUL SENTONGO	M23M19/256	B00237
NANKYA MARGARET PENINAH	M23M10/255	B00326

Proposal for Predicting Employee Attrition

Problem Definition

The rate of employee turnover in Uganda is a significant concern, with various factors contributing to high attrition rates. According to the Uganda Bureau of Statistics' National Labour Force Survey 2021, there is a notable movement of employees between jobs, which impacts various sectors differently (UBOS, 2021)

Human Resource Managers in Uganda have highlighted that this high turnover rate is challenging for companies, as it affects their ability to invest in and retain skilled employees. The frequent job changes among employees mean that companies face increased costs related to recruitment, training, and development, which are not always recuperated before employees leave for other opportunities (Monitor, 2021)

Efforts to address these issues include strategies such as conducting exit interviews to understand why employees leave, implementing stay interviews to identify potential issues while employees are still with the company, and aligning talent with appropriate job roles to enhance job satisfaction and retention (Monitor, 2021)

Our goal is to develop a predictive model to identify employees who are likely to quit. This will enable the HR team to proactively address potential attrition issues.

Data Collection

The dataset used includes various attributes of employees such as Job Involvement, Education, Job Satisfaction, Performance Rating, Relationship Satisfaction, and Work-Life Balance. This data is sourced from a public dataset available on Kaggle. We acknowledge that this dataset is from a different geographical location and may require contextual adaptation for our specific scenario in Uganda.

Data Preprocessing

Data preprocessing is critical to ensure the quality and reliability of the dataset. Steps involved include handling missing values, encoding categorical variables, normalizing

numerical features, and addressing any outliers. Additionally, we will perform data cleaning to ensure consistency and accuracy in the dataset.

Exploratory Data Analysis (EDA)

EDA will involve visualizing the distribution of various features and their relationships with the target variable (attrition). Techniques such as histograms, box plots, and correlation matrices will be used to gain insights into the data. Summary statistics will help identify key trends and patterns that may influence employee attrition.

Feature Engineering

Feature engineering involves creating new features or modifying existing ones to improve the predictive power of the model. For example, we may derive new features such as tenure, number of promotions, and average hours worked per week. Interaction features that capture the combined effect of multiple variables may also be created.

Model Selection

Several machine learning algorithms will be considered, including Logistic Regression, Decision Trees, Random Forests, and Gradient Boosting Machines (GBM). Additionally, a deep learning model will be built to address the task. Each algorithm has its strengths and will be evaluated based on its performance metrics.

Model Training

The selected models will be trained on the preprocessed dataset. Techniques such as cross-validation will be employed to ensure the robustness of the models. Hyperparameter tuning will be conducted using methods like Grid Search or Random Search to optimize model performance.

Model Evaluation

Models will be evaluated using metrics such as accuracy, precision, recall, F1-score, and the Area Under the Receiver Operating Characteristic Curve (AUC-ROC). The model with the best performance on these metrics will be selected for deployment.

Model Deployment

The final model will be deployed in a production environment where it can be used to predict employee attrition on new data. This will involve setting up an API or integrating the model into the existing HR management system.

Monitoring and Maintenance

Once deployed, the model will be continuously monitored to ensure its performance remains consistent over time. Regular updates and retraining may be required to adapt to changes in employee behavior and organizational dynamics.

Documentation and Reporting

Comprehensive documentation will be maintained throughout the project, detailing each step from problem definition to model deployment. Reports will include findings from the EDA, model performance metrics, and insights derived from the predictive model.

Feedback and Iteration

Feedback from stakeholders will be gathered to refine and improve the model. Iterative cycles of testing and enhancement will be conducted to ensure the model remains effective and aligned with business objectives.

By leveraging machine learning methodologies, this project aims to provide actionable insights that can help the HR team in Uganda mitigate employee attrition, ultimately saving costs and improving organizational efficiency.

References

Monitor, 2021. *Monitor Publications.* [Online]
Available at: https://www.monitor.co.ug/uganda/magazines/jobs-and-career/hr-managers-grapple-with-high-turnover-rate-1475898

UBOS, 2021. *National Labour Force Survey,* s.l.: s.n.

Toggl Blog. (n.d.). Cost of Hiring an Employee. Retrieved from https://toggl.com/blog/cost-of-hiring-an-employee

Kaggle. (n.d.). IBM HR Analytics Employee Attrition & Performance. Retrieved from

https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset