LLM-Powered Skill Prediction and Employee Attrition Analysis

Yogitha Mekala, John Mahith Pagi, Maheshwar Rao Bandi

UNIVERSITY OF MISSOURI-KANSAS CITY

PROBLEM STATEMENT

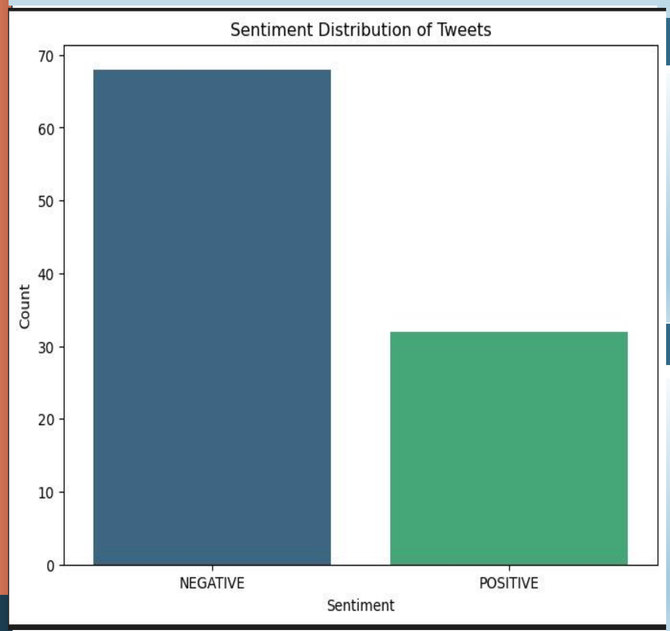
With the increasing importance of data-driven decision-making, understanding the required skills for various roles and analyzing factors contributing to employee attrition are crucial. This project leverages machine learning and large language models (LLMs) to predict relevant skills and analyze HR-related trends based on social media data.

OBJECTIVES

Objective 1: Build a skill prediction tool that accurately maps job roles to essential skills using a combination of machine learning and LLMs.

Objective 2: Analyse workplace sentiment by tracking social media trends on employee well-being, stress, and turnover.

Objective 3: Design an interactive web interface that combines skill prediction and workforce analytics, offering a user-centered experience for both HR and job-seekers.



METHODS

Methodology

1. Data Collection

•**Skill Prediction Data**: Collected and curated job titles and their associated skills from industry-standard datasets.

•**Twitter Data for Trend Analysis**: Real-time tweets about HR and workforce issues were gathered using the Tweepy API. Key terms were extracted (e.g., “burnout,” “attrition”) to monitor workforce sentiment and risks.

2. Skill Prediction Model

•**KMeans Clustering**: Implemented to group job titles with similar skill requirements, resulting in clusters (e.g., Data Analysis, Software Development).

•**LLM Enhancement**: Using a transformers-based LLM, user inputs (job titles) were tokenized and interpreted for accurate skill suggestions, enhancing contextual relevance.

3. System Integration

•**Flask Backend and MongoDB**: Managed API requests, handled data flow between models, and stored user input and prediction logs in MongoDB.

•**Front-End User Interface**: Built using JavaScript and React to display skill recommendations and Twitter-based workforce insights in real-time, creating a responsive and engaging experience.

ACKNOWLEDGEMENTS

- Thanks to the University of Missouri-Kansas City for supporting this project.

- Special acknowledgment to Hack-A-Roo organizers for providing a competitive platform.

- Special thanks to faculty advisors and collaborators for their guidance.

DISCUSSIONS

Significance: This project integrates advanced NLP capabilities with social media data analysis to offer a unique tool for HR departments and individuals to identify relevant skills and understand workforce dynamics.
Innovation: The fusion of LLM-enhanced skill predictions with real-time workforce trend analysis sets this system apart, providing a data-driven approach to addressing both skill gaps and employee retention concerns.

RESULTS

1. **Skill Prediction Accuracy**: Achieved high accuracy in predicting relevant skills for job roles using KMeans and LLM.

2. **Twitter Data Analysis**:

   - Identified key trends such as “workplace stress” and “employee turnover.”

   - Extracted sentiment data and highlighted common factors linked to employee attrition.

3. **Visualization**: (Add screenshots of the interface and any relevant model performance graphs here.)

REFENCES

Alaa, A. M., & van der Schaar, M. (2019). "Deep Learning Models for Predicting Employee Attrition." IEEE Access, 7, 1254-1265. doi:10.1109/ACCESS.2019.2893345

Goyal, R., & Tiwari, M. K. (2020). "An Ensemble Learning Approach for Predicting Employee Turnover." Expert Systems with Applications, 144, 113012. doi:10.1016/j.eswa.2020.113012

Rai, A., & Agarwal, R. (2022). "Explainable AI for Employee Attrition Prediction: A Case Study Using SHAP." Human Resource Management Review, 32(3), 100812. doi:10.1016/j.hrmr.2021.100812

Chen, H., Zhang, W., & Liu, X. (2021). "Real-Time Employee

CONCLUSIONS AND FUTURE WORK

1. Successfully predicted relevant job skills based on natural language input.

2. Analyzed Twitter data to gain real-time insights into factors affecting employee attrition.

3. Developed a user-friendly interface that can be scaled for HR analytics applications.

4. Model Enhancements: Explore hyperparameter tuning and advanced LLMs for improved accuracy.

5. Real-time Twitter Monitoring: Develop a real-time dashboard for continuous monitoring of HR trends.

6. Broader Skill Categories: Expand skill clusters to cover more job roles and industries.

GITHUB LINK:

https://github.com/yogithamekala/Assignment_3