Detailed Report of LinkedIn Job Postings Analysis

Introduction

This report summarizes the findings from an analysis of LinkedIn job postings dataset. The dataset includes company details, job titles, descriptions, and salary information.

Data Overview

The dataset comprises job-related information from LinkedIn job postings, including key features like company details, job-specific information like titles, descriptions, and salary ranges.

Data Cleaning

Handled Missing Values: Removed rows with missing company names and job descriptions. Created a binary indicator for missing values in the `domain` field.
Data Type Correction: Verified and corrected data types for all columns.
Outlier Treatment: Capped salary and follower count outliers at the 95th percentile.
Text Data Preprocessing: Cleaned job descriptions by removing special characters.

Exploratory Data Analysis (EDA)

Descriptive Statistics: Analyzed company sizes, salary ranges, and job posting views.
Correlation Analysis: Found low correlations between company size and salaries, and job views with other variables.
Category Analysis: Identified common job titles, frequent hiring companies, and the distribution of experience levels.

Feature Engineering

Created Salary Brackets: Categorized salaries into 'Low', 'Medium', and 'High' brackets.
Encoded Categorical Data: Used label encoding and frequency encoding for categorical data like experience levels, job titles, and company names.

Predictive Modeling

Random Forest Classifier: Developed to predict salary brackets. Showed reasonable performance with better accuracy in identifying 'Low' salary bracket.
Logistic Regression: A simpler model with lower accuracy, showing significant bias towards the 'Low' salary bracket.

Data Visualization

Salary Distributions: Visualized the distribution of minimum, median, and maximum salaries.
Job Posting Frequency by Company: Showed the frequency of job postings for top companies.
Location-Based Analysis: Analyzed the distribution of job postings across top locations.

Conclusions and Recommendations

- The analysis provided valuable insights into the job market trends on LinkedIn, including salary distributions, company hiring frequencies, and popular job roles.
- The predictive models, especially the Random Forest Classifier, can be refined further for enhanced accuracy and reduced bias.
- Incorporating additional data, such as temporal information, could enable more comprehensive analyses.

Future Work

- Iteration on the predictive models with more advanced techniques or additional data.
- Exploration of other aspects of the dataset, such as deeper text analysis of job descriptions.
- Continuous update and refinement of the analysis with new data.