

Fraudulent Job Posting Detection – EDA Summary Report

1. Introduction

This report summarizes the Exploratory Data Analysis (EDA) performed on the Fraudulent Job Posting dataset. The objective is to understand patterns and indicators that distinguish real and fraudulent postings.

2. Data Cleaning Steps

- Dropped columns with over 60% missing values.
- Filled remaining missing text fields with 'Unknown' and categories with 'Not Specified'.
- Converted the target column to fraud_label (0 = Real, 1 = Fraud).

3. Univariate Analysis

- Countplots explored distributions of job roles, industries, employment types, and experience levels.

4. Bivariate & Multivariate Analysis

- Fraud rate comparisons show higher fraud among telecommuting jobs and jobs without requirements.
- Correlation and cluster heatmaps highlighted relationships between numeric features.
- Barplots showed clear clustering patterns where fraud postings group together.

5. Key Insights

- Remote job postings show increased fraud likelihood.
- Multivariate patterns suggest strong combined indicators of fraud.
- Requirements often blank.

6. Recommendations

- Flag postings missing company details or proper descriptions.
- Add stricter verification for telecommuting roles.
- Use automated rules leveraging text length and missing field patterns.