

FAKE JOB DETECTION PROJECT REPORT

MACHINE LEARNING & EXCEL DASHBOARD PROJECT

PROJECT REPORT

**BY
MEVINA P.D**

DATA ANALYST

**UNDER THE GUIDANCE & SUPERVISION OF
Miss Amrutha**

Palakkad , kerala – 678702



**TECHOLAS
TECHNOLOGIES**

CONTENTS

INTRODUCTION

ABSTRACT

OBJECTIVES

METHODOLOGY

DATA ANALYSIS

MODEL DEVELOPMENT & RESULTS

DASHBOARD OVERVIEW

CONCLUSION & FUTURE SCOPE

INTRODUCTION

With the growth of online job portals and digital recruitment platforms, job searching has become easier and more accessible. However, this has also increased fake job postings and online recruitment fraud. These fraudulent advertisements mislead job seekers and may cause financial loss, emotional stress, and misuse of personal information.

Fake job postings often contain misleading details, incomplete company information, and suspicious language. Due to the large number of job advertisements posted daily, manually identifying fake jobs is difficult and time-consuming.

This project focuses on developing a machine learning-based system to automatically detect fake job postings. The system analyzes job descriptions, company details, and requirements to classify postings as genuine or fake. By using data preprocessing, exploratory data analysis, and machine learning models, this project aims to improve fraud detection accuracy and create a safer online job search environment.

ABSTRACT

This project presents a machine learning-based approach to detect fake job postings. The dataset contains job descriptions, company information, requirements, and other attributes.

Exploratory Data Analysis (EDA) and preprocessing techniques were applied to clean and transform the data. Multiple classification models were trained and evaluated to select the best-performing model.

OBJECTIVES

- To analyze job posting data using EDA
- To clean and preprocess textual and numerical data
- To build machine learning models for fake job detection
- To compare model performance
- To identify the best model
- To improve online job security

METHODOLOGY

Step 1: Data Collection

The dataset was collected from kaggle.

Step 2: Data Preprocessing

- Handling missing values
- Text cleaning
- Tokenization
- Vectorization (TF-IDF)
- Encoding categorical variables

Step 3: Exploratory Data Analysis

- Fake vs real job distribution
- Feature correlation
- Text length analysis

Step 4: Model Building

- Logistic Regression
- Naive Bayes
- Decision Tree
- Random Forest
- Support Vector Machine

Step 5: Model Evaluation

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

DATA ANALYSIS

Data analysis was conducted to understand the structure and quality of the job posting dataset. The dataset included both genuine and fake job listings with features such as job title, company details, job description, location, and salary information.

Missing values and inconsistent data were identified and handled during preprocessing. Textual data was analyzed to study keyword patterns, description length, and frequency of suspicious terms. Categorical features were also examined to understand their distribution.

Exploratory Data Analysis (EDA) was performed using charts and graphs to identify patterns and trends in the data. The analysis showed that fake job postings often had incomplete information, unusual salary patterns, and misleading keywords, while genuine postings were more detailed and reliable.

These findings helped in selecting relevant features and improving the performance of machine learning models.

MODEL EVALUATION

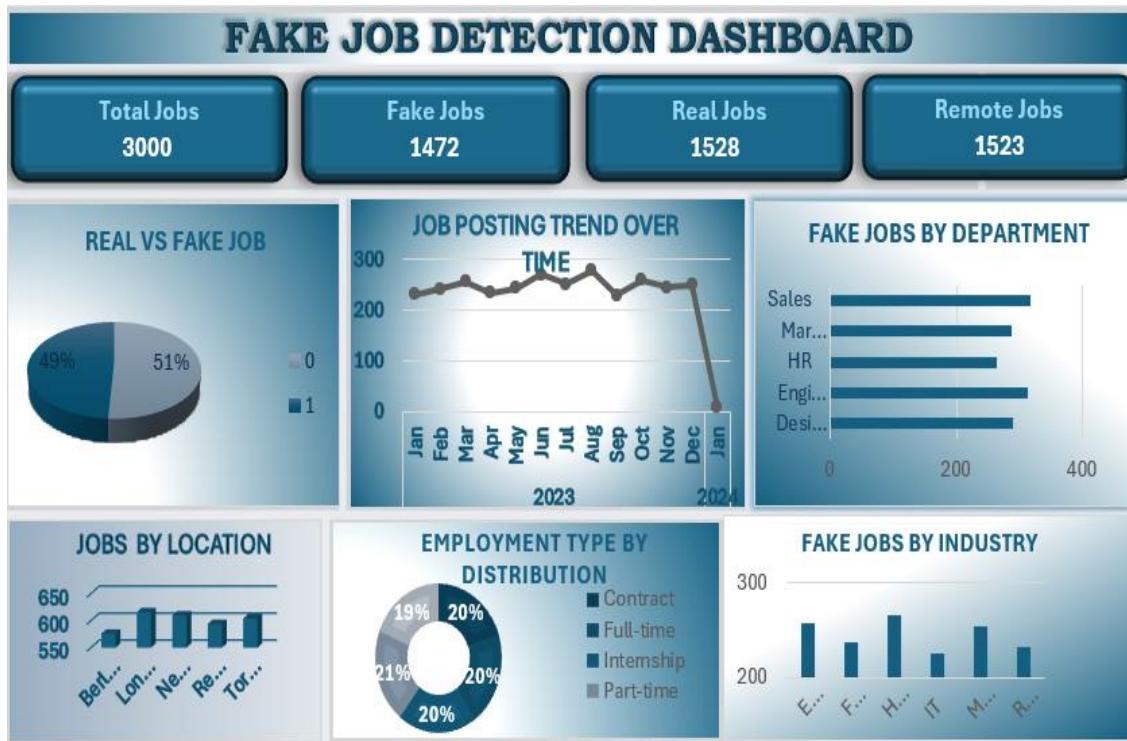
Model Performance Comparison:

Logistic Regression	Good
Decision Tree	Good
SVM	Very good
Random Forest	Best

Best Model: Random Forest Classifier

- Highest accuracy
- Reduced overfitting
- Better generalization

DASHBOARD OVERVIEW



- Displays key indicators such as total jobs, fake jobs, real jobs, and remote jobs.
- Shows the distribution of real and fake job postings using charts.
- Presents job posting trends over time for better analysis.
- Highlights fake jobs by department and industry.
- Analyzes job postings by location and employment type.
- Helps in identifying fraud patterns and making informed decisions.

CONCLUSION & FUTURE SCOPE

Conclusion

This project successfully demonstrates the application of machine learning in detecting fake job postings. The developed system improves online job safety and helps users avoid fraud.

Future Scope

- Use larger real-time datasets
- Apply deep learning models
- Integrate browser extensions
- Develop mobile applications
- Improve NLP techniques