

# Resume Classification Using Machine Learning and Deep Learning Models

**Objective:** To build an intelligent system capable of automatically classifying resumes into predefined job categories such as Data Science, HR, DevOps, etc., using both traditional machine learning models and deep learning (RNN) techniques.

## Project Overview:

### 1. Data Preprocessing:

Collected a labeled dataset of resumes with associated job categories.

Cleaned and preprocessed the resume text data through custom NLP functions (e.g., tokenization, stopwords removal, lemmatization, etc.).

Encoded target categories using label encoding for classification.

### 2. Feature Engineering:

Implemented **TF-IDF (Term Frequency-Inverse Document Frequency)** vectorization to convert raw text into numerical features for ML models.

For deep learning, used **Keras Tokenizer** and **sequence padding** to prepare input data for LSTM layers.

### 3. Model Training:

> Trained **traditional ML classifiers**:

`KNeighborsClassifier` wrapped in `OneVsRestClassifier`

`Multinomial Naive Bayes (MNB)`

> Achieved **~98.3% accuracy** with KNN and MNB models.

> Trained a **Recurrent Neural Network (RNN)** using Keras:

Model architecture: `Embedding` → `SpatialDropout1D` → `LSTM` → `Dense` → `Softmax`

Achieved **~98.34% accuracy** on test set.

### 4. Evaluation:

Compared models using metrics such as **accuracy, precision, recall, and F1-score**.

Observed high performance across all models, with robust classification across 25 different job categories.

### 5. Deployment Preparation:

> Used `pickle` to **serialize and save**:

TF-IDF vectorizer

Trained ML models (KNN and MNB)

Trained RNN model using Keras

> Developed **prediction pipelines** for real-world resume input using all saved models.

### 6. Testing & Prediction:

Passed a sample resume (`resume_1`) through the complete pipeline.

Preprocessed the resume, transformed it into features, and predicted its category.

All models successfully predicted the correct category: **Data Science**.

## Tools & Technologies Used:

**Languages:** Python

**Libraries:** scikit-learn, Keras, TensorFlow, NLTK, NumPy, pandas, pickle

**Techniques:** NLP, TF-IDF, Multi-class classification, LSTM, Tokenization, Sequence Padding

## Key Achievements:

Built a full-stack machine learning system for multi-class text classification.

Demonstrated strong performance from both classical and deep learning models.

Enabled real-time resume classification with deployable models.

Validated performance using classification metrics and sample resume prediction.