

Arshid T M

DATA SCIENTIST

📍 Kannur, Kerala ✉ arshidtm2001@gmail.com ☎ +91 8078196576 < / > Leetcode in Arshid T M
🔗 Arshidtm

Profile Summary

Data Science and Machine Learning enthusiast with expertise in building predictive models, deep learning, and NLP. Skilled in data preprocessing, feature engineering, and EDA. Proficient in Python, TensorFlow, and Scikit-Learn for model development and deployment. Passionate about leveraging AI to solve real-world problems and drive business growth.

Skills

- **Programming Language:** Python
- **Data Analysis & Visualization:** Pandas, NumPy, Matplotlib, Seaborn, Tableau
- **Database Management:** PostgreSQL
- **Machine Learning:**
 - **Regression:** Linear Regression, Random Forest Regressor
 - **Classification:** KNN, Naive Bayes, SVM, Decision Trees, Random Forest Classifier
 - **Ensemble Learning:** Bagging, XgBoost
 - **Clustering:** K-Means, DBSCAN
 - **Advanced Topics:** Natural Language Processing (NLP), Deep Learning

Projects

Sentiment Analysis

[GitHub](#)

Developed a **Machine learning**-based sentiment analysis project that applies **NLP** techniques like tokenization, stemming, and vectorization (TF-IDF & Word2Vec). Trained using **Random Forest and Multinomial Naïve Bayes**, achieving **96%** accuracy with Grid Search tuning. Developed as a web application using **Flask** for real-time predictions.

- **NLP Preprocessing:** Tokenization, stemming, vectorization (TF-IDF & Word2Vec)
- **Machine Learning Models:** Implemented Random Forest & Multinomial Naïve Bayes for text classification.
- **Hyperparameter Tuning:** Used Grid Search to optimize model performance.
- **Deployment:** Hosted as a Flask web application for real-time predictions.
- **Tools Used:** Python — Pandas — NumPy — Scikit-Learn — NLTK — Matplotlib — Seaborn — Flask

Movie Recommendation System – Content-Based Filtering

[GitHub](#)

Developed a **content-based movie recommendation system** that suggests movies based on the director, actors, and genre. Applied text preprocessing using **NLP** techniques, vectorized movie data using **Bag of Words (BoW)**, and calculated similarity scores using **cosine similarity** for accurate recommendations. Developed as an interactive **Streamlit** web app for seamless user experience.

- **NLP Preprocessing:** Cleaned and processed movie metadata
- **Vectorization:** Used **Bag of words (BoW)** for the representation of characteristics.
- **Similarity Computation:** Implemented **cosine similarity** for movie recommendations.
- **Deployment:** Built and deployed using **Streamlit** for a user-friendly interface.
- **Tools Used:** Python — Pandas — NumPy — Scikit-Learn — NLTK — Streamlit

Mini Project

WhatsApp Chat Analyzer

[GitHub](#)

- Developed a WhatsApp Chat Analyzer that extracts insights from exported .txt files, performing EDA to visualize message frequency, participant activity, and emoji usage. Applied stop-word removal for text preprocessing and deployed as a Streamlit Web app for real-time analysis.
- Extracts and analyzes WhatsApp chat data from text files.
- Identifies chat patterns, word frequency, and emoji usage.
- Applied text preprocessing and stopwords removal for cleaner analysis.
- Built using Streamlit for an interactive user experience.
- **Tools Used:** Python — Pandas — Matplotlib — Seaborn — Streamlit

HR Analytics Dashboard

[Tableau](#)

- Developed an interactive HR Analytics Dashboard using Tableau to analyze key workforce metrics, including employee count, attrition rate, department-wise attrition, job satisfaction, and education-wise attrition.
- Visualizes attrition trends by department, education field, and gender for data-driven HR decision-making.
- Analyzes the age distribution of the employees and the satisfaction ratings in various job roles.
- Built using Tableau for an interactive and user-friendly experience.
- **Tools Used:** Tableau

Email Spam Detection

[GitHub](#)

- Developed an Email Spam Detection System using NLP and machine learning to classify emails as spam or ham.
- Applied text preprocessing techniques like tokenization, stopwords removal, and stemming for data cleaning.
- Identified frequent words in spam and ham emails using WordCloud visualization.
- Used TF-IDF vectorization to convert text into numerical features.
- Trained and compared multiple models (GaussianNB, MultinomialNB, BernoulliNB), with TF-IDF + MultinomialNB achieving the best precision score.
- Streamlit Web App Deployed for Real-Time Email Classification.
- **Tools Used:** Python — Pandas — Scikit-learn — NLTK — Streamlit

Education

Data Science

2024 - present

Prototype, Remote

Master of Science in Data Science and Business Analysis

2022 - 2024

Rathinam College of Arts and Science - Bharatiyar University, Coimbatore

CGPA: 7.9

Bachelor of Science in Mathematics

2019 - 2022

Sir Syed College - Kannur University, Kannur

CGPA: 7.067