DHARUN B

Aspiring Data Scientist

9791817437 | bdharun1912@gmail.com | Coimbatore, Tamil Nadu, India | LinkedIn | Portfolio

SUMMARY

As a recent graduate with a solid background in data science, I am passionate about utilizing data to facilitate impactful business decisions. My academic pursuits have honed my analytical abilities and provided me with hands-on experience in machine learning, statistical analysis, and data visualization. I am eager to bring my skills and enthusiasm to your team, contributing to innovative data-driven solutions.

WORK EXPERIENCE

Data Scientist Intern - Qtree Technologies, Coimbatore **Jul 2024 - Present**

Key Responsibilities:

- Gained hands-on experience in data manipulation, data analysis, and statistical modeling using Python and SQL
- Developed machine learning models, including regression, classification, and clustering techniques.
- Performed data preprocessing tasks such as data cleaning, feature engineering, and handling missing values.
- Built and deployed interactive dashboards for data visualization using **Seaborn and Matplotlib**.

Key Achievements:

- Created a predictive model for **Cyberbullying detection**, achieving an accuracy of **90%** using **[ML techniques like LinearSVC, TF-IDF vectorizer]**.
- Developed a data pipeline to automate the collection, cleaning, and analysis of data, reducing processing time by **65%**.
- Presented a final capstone project on Cyberbullying, demonstrating proficiency in machine learning algorithms and data storytelling.

SKILLS

Programming Language	**:**	Python
Libraries and Frameworks	**:**	Pandas, NumPy, Scikit-learn, Xgboost, Adaboost, SciPy, Beautifulsoup, Streamlit
Visualization	**:**	Matplotlib, Seaborn, Plotly, Altair
Tools	**:**	SQL, Git, Jupyter Notebook

PROJECTS

Cyberbullying Detection System

- Implemented data preprocessing techniques, including stopwords removal and text vectorization using TF-IDF.
- Trained and optimized a **LinearSVC** model for text classification to detect bullying content.
- Integrated a pre-fitted TF-IDF Vectorizer and a machine learning model to make accurate predictions.
- The system flags potentially harmful content, aimed at improving online safety and preventing cyberbullying.
- **Technologies:** Python, Streamlit, Scikit-learn, Natural Language Processing (NLP), TF-IDF Vectorizer, LinearSVC.

Sentiment Analysis on Restaurant review

- Built an NLP model to classify restaurant reviews as positive or negative using **CountVectorizer** and **Textblob.**
- Preprocessed text data: cleaning, tokenization, stopword removal, and stemming.
- Applied **Logistic Regression**, **Naive Bayes**, and **XGBoost** for sentiment prediction.
- Achieved 95**% accuracy** and deployed the model via **Streamlit** for real-time sentiment analysis.
- **Technologies**: Python, NLTK, Scikit-learn, XGBoost, Streamlit.

Brest cancer classification

- Built a web application using **Streamlit** to predict breast cancer outcomes (Benign or Malignant) based on patient data.
- Implemented a **Logistic Regression** model to classify tumors by analyzing 30 clinical features from a patient dataset.
- Enabled users to enter **Patient ID** to retrieve and display patient-specific data and generate real-time predictions.
- Used **joblib** for model persistence and integrated data preprocessing with **Pandas** for clean input handling.
- **Technologies**: Python, Streamlit, Logistic Regression, scikit-learn, Pandas, NumPy, joblib.

EDUCATION

Sri Ramakrishna College of Arts and Science,Coimbatore - Bsc,Information Technology -*GPA:7.4* **Jan 2021 - Apr 2024**

Relevant Coursework : Data Science, Machine Learning, Python Programming language, Data Visualization

CERTIFICATIONS

Python for Data Science – IBM Developer skills **Mar 2024**

British Airways - Data Science Job Simulation - Forage **Aug 2024**

Python (Basics) - HackerRank **Sep 2024**