Bharathkumar M S

Address: Melukote, Karnataka
Phone No: +91 6362403176
Email ID: bharathkumarms3011@gmail.com
LinkedIn: https://www.linkedin.com/in/bharathkumar-m-s

GitHub profile: Link
Portfolio: Link
Medium : Link
Hackerrank: Link

CAREER OBJECTIVE

A highly motivated and analytical individual with a passion for data science and machine learning with good knowledge of gathering, cleaning, organizing and analyzing data using Python along with good understanding of Statistics, SQL, Machine Learning Algorithms. Seeking a challenging role as a data scientist to utilize my skills and knowledge in solving real-world problems.

ACADEMIC QUALIFICATION

Qualification	Institution	University/Board	Year of Passing	Percentage
B.E (Mechanical)	Malnad College of Engineering, Hassan (Autonomous)	VTU	2022	8.29 CGPA
2nd PUC (PCMB)	Anikethana PU college, Mandya	PU Board	2018	87.5
SSLC	Morarji Desai Residential School, Mandya	KSEEB	2016	89.4

TECHNICAL SKILLS

- **Programming Languages**: Python, SQL.
- **Python Libraries and Frameworks**: Pandas, Numpy, Matplotlib, Seaborn, Scikit-learn, Tensorflow, Keras.
- **Machine Learning**: Linear regression, Logistic regression, Decision Tree, Random Forest, SVM, K-means clustering, KNN, Naive Bayes, PCA, Gradient boosting, XG Boosting, Recommendation Systems, EDA, Pre-processing, Feature Engineering, Model building, Deployment.
- **Deep Learning**: Artificial Neural Network(ANN), Convolution Neural Network(CNN), Computer vision, NLP, RNN, LSTM, Encoder, Decoder, GRU, BOW, TFIDF, Embedding, Text Preprocessing.
- **Mathematics**: Statistics, Probability, Linear algebra, calculus.
- **Web Development**: Flask, Streamlit, HTML.
- **Databases**: MySQL.
- **Tools & Technologies**: Power BI, Docker, MLOps, MLflow, CICD, Github Actions, Git & Github.
- **IDE**: PyCharm, Visual Studio, Jupyter Notebook, Google COLAB.

EXPERIENCE

DATA SCIENCE INTERN at Innomatics Research Labs (Link) **[02/2023– 06/2023]**

- Worked on some tasks related to python and Flask.
- Worked on some machine learning projects.
- Worked on Plant Disease Classification Project.

DATA SCIENCE INTERN at iNeuron.ai (Link) **[10/2022 – 02/2023]**

- Worked with a variety of tools and techniques, including Python, SQL, Pandas, Numpy, Matplotlib and scikit-learn.
- Analyse the data by visualisation using seaborn and matplotlib.
- Worked on End to End project on flight fare prediction.

DATA SCIENCE INTERN at Exposys Data Labs (Link) **[12/2021 - 02/2022]**

- Worked on various data cleaning and data preprocessing techniques.
- Performed Data cleaning, Exploratory Data analysis and Feature engineering Techniques.
- Worked on the Healthcare Diabetic Prediction Project.

PROJECTS

FLIGHT FARE PREDICTION [web app Link]

- **Project Objective**: Developed a flight fare prediction system to estimate airfare prices based on various input parameters and route details.
- **Data Collection**: Gathered a comprehensive dataset containing historical flight booking information, including departure and arrival cities, dates, airlines, and ticket prices, using Pandas for data manipulation and storage.
- **Data Preprocessing**: Conducted extensive data preprocessing, handling missing values, and encoding categorical variables to prepare the data for machine learning.
- **Feature Engineering**: Transformed data by engineering new features, including "Journey Day" and "Journey Month" from the "Date of Journey" column, as well as other relevant features, to enhance the predictive accuracy of the model.
- **Exploratory Data Analysis (EDA)**: Performed in-depth EDA using data visualization libraries like Matplotlib and Seaborn in conjunction with Pandas to gain insights into the dataset and identify trends and patterns. Utilized Power BI for advanced visualizations.
- **Model Selection**: Explored a range of machine learning algorithms, including regression models like Linear Regression, Decision Trees, and Random Forest. Leveraged XGBoost to achieve an impressive accuracy of 85%.
- **Hyperparameter Tuning**: Fine-tuned model hyperparameters using technique like Grid Search to optimize predictive accuracy.
- **Cross-Validation**: Implemented cross-validation strategies to assess the model's generalization performance and mitigate overfitting.
- **Evaluation Metrics**: Evaluated model performance using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) to quantify prediction accuracy.
- **Docker Containerization**: Utilized Docker to containerize the entire project, ensuring consistent and portable execution across different environments.
- **MLflow Model Tracking**: Implemented MLflow for model tracking and management, enabling version control and reproducibility of model experiments.
- **Web Application**: Designed and developed an interactive web application using Flask and Streamlit, enabling users to input flight details and receive real-time fare predictions.
- **Deployment**: Successfully deployed the Flight Fare Prediction system as a web application, providing travelers with a user-friendly tool for estimating ticket prices.

SMS SPAM CLASSIFIER [web app Link]

- **Project Objective**: Developed an SMS spam classifier to automatically identify and filter out spam messages from legitimate ones.
- **Data Collection**: Gathered a diverse dataset containing both spam and non-spam (ham) SMS messages for training and evaluation.
- **Data Preprocessing**: Conducted data preprocessing tasks, including text cleaning, tokenization, and removing stop words, to prepare the text data for machine learning.
- **Feature Extraction**: Utilized techniques like Count Vectorization and TF-IDF to convert text messages into numerical feature vectors for machine learning models.
- **Model Selection**: Explored various machine learning algorithms, such as Naive Bayes and Logistic Regression, to identify the most suitable model for spam classification.
- **Hyperparameter Tuning**: Fine-tuned model hyperparameters and performed cross-validation to optimize the classifier's performance.
- **Evaluation Metrics**: Evaluated model performance using metrics such as precision, recall, F1-score, and accuracy to assess its effectiveness in correctly identifying spam messages.
- **Web Application**: Designed and developed a user-friendly web application using Streamlit, allowing users to input SMS messages and receive spam predictions in real-time.
- **Deployment**: Successfully deployed the SMS spam classifier as a web app, making it accessible to users for practical use.

AGRICULTURE PLANT DISEASE CLASSIFIER [web app Link]

- **Project Objective**: Developed a CNN-based plant disease classifier model to classify plant diseases based on leaf images, aiding farmers in early disease detection and crop protection.
- **Data Collection**: Collected a diverse dataset comprising images of healthy and diseased plants, encompassing various plant species and diseases, to train and evaluate the classification model.
- **Data Preprocessing**: Conducted extensive data preprocessing tasks, including resizing, normalization, and augmentation of images, to prepare the dataset for training and minimize overfitting.
- **Deep Learning Architecture**: Designed and implemented a Convolutional Neural Network (CNN) architecture using popular frameworks like TensorFlow and Keras to extract and learn relevant features from plant images.
- **Model Training**: Trained the CNN model on the preprocessed dataset, employing techniques such as transfer learning using pretrained models like VGG, ResNet, Inception and EfficientNet to enhance classification performance.
- **Evaluation Metrics**: Assessed model performance using evaluation metrics such as accuracy, precision, recall, and F1-score, ensuring reliable disease classification results.
- **Web Application**: Created an interactive web application using Streamlit, allowing users to upload plant images and receive instant disease classification results.
- **Deployment**: Successfully deployed the Plant Disease Classification system as a web app, providing farmers and agriculture professionals with a practical tool for disease identification.
- **User Interface**: Designed an intuitive user interface with image upload functionality, facilitating the automatic identification and classification of plant diseases from images.

CERTIFICATIONS

- **DATA SCIENCE CERTIFICATION** by Besant Technologies [Link]
- **PYTHON CERTIFICATION** by Besant Technologies [Link]
- **FULL STACK DATA SCIENCE COURSE** by Ineuron.ai [Link]
- **INTRODUCTION TO MACHINE LEARNING** by Coursera [Link]
- **DATA SCIENCE WITH PYTHON** by simplilearn [Link]

EXTRACURRICULAR ACTIVITIES

- NSS volunteer

PERSONAL SKILLS

- Motivating and goal oriented
- Willingness to learn and ability to be a good team player
- Quick Learner
- Problem Solving
- Good analytical and logical skills
- Having leadership quality
- Having good convincing power
- Ability to tackle critical situation

LANGUAGES

English

Professional Working Proficiency

Kannada

Native or Bilingual Proficiency

HOBBIES

Listening to music	**Watching movies**	**Playing cricket**	**Cycling**