# Sowjanya Bojja

Data Scientist

**+91 8073226086**
simplysowj@gmail.com
https://github.com/simplysowj
**https://www.linkedin.com/in/sowjanya-bojja/**

## CERTIFICATIONS:

Certified IBM full stack data scientist from Innomatics Research labs(one of the top growing Edu-Tech Company in Hyderabad)

Certificate of training in SQL,Tableau,Advanced Excel,ML and DL in Internshala Training.

## Education:

MSC in Data Science(Pursuing),
Btech in Electronics and Communication Engg.

## EXPERIENCE

July 2022 - November 2022

Data Science Intern in Innomatics Research Labs | Hyderabad, India

Projects:

Quora Duplicate question detector (capstone project)

Tasks Developed :

application using Streamlit

Applied machine learning algorithms

Experiment Tracking and Model Management using MLFLOW

Different vectorization techniques including bert Workflow orchestration with Prefect

Natural Language Processing

Cloud Deployment Heroku

GitHub Projects:

1)NLP project using Streamlit for sentimental analysis on the IMDB Movie reviews tweets dataset,

here in this project I used vectorization technique like #bert to convert

## SKILLS

**Technical Skills:**

- **Programming Languages:** Python, Java, C

- **Web Development:** HTML, CSS, JavaScript, Bootstrap

- **JavaScript Libraries & Frameworks:** Node.js, React.js

Java Framework:SpringBoot

- **Microservices & Containers:** Docker

- **Data Visualization:** Tableau, Excel

- **Web Frameworks:** Flask, Streamlit

- **Machine Learning & Deep Learning:** (ML, DL), NLP

NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, NLTK

ML: Data Exploration & Analysis, Data Modelling, Statistics and Probability, Linear Regression, Gradient Descent, Logistic Regression, Regularization, SVM, KNN, Decision Trees, Random Forest, Ensemble Techniques, Bagging & Boosting, Cross-Validation, Cluster Analysis,Tensorflow,SKlearn, Hyper parameter tuning,Experiment tracking

text data into numerical vectors

2) covid - 19 twitter sentiment analysis with streamlit application using python.

3) MLOps task-Diamond Dataset ML pipeline orchestration and experiment tracking

4) Experiment Tracking and Model Management for Diamond price prediction using MLFlow

5) streamlit app for Diamonds dataset EDA and ML Model

6) Machine learning project - Regression Problem Title: Concrete compressive strength

7) Python Flask CRUD Web Application project 8) Portfolio Application Project

9) Python Web Scraping Project with EDA

10) Open Pub Application using Streamlit python

Education Project:

IPL Data Analysis using SQL(Postgresql) and Data visualization using Excel.

Analysis of Fitbit fitness tracker data to deliver marketing and business solutions to WeFit and its subsidiaries

Tableau Project

Portfolio using React JS

Minor project:

Exam seating arrangement system using JAVA with GUI Swing.

Industry Experience:

Plsql Developer Tools: TOAD 7.6.0.11, PL/SQL Developer, Forms 61, Reports 6i

Company TCS

Client Electronic Arts-India Location Bangalore, India (2010 to 2014) (Nov 2010 May2014)

Client Tata TeleServices Limited - India (2008 to 2010) Mar 2008-Oct 2010

and model management using MLflow. MLOPS,Workflow orchestration with Prefect 2.4

Cloud Deployment : Heroku GitHub

NLP: Tokenization, Bag of words, Stemming, Lemmatization, POS Tagging, TFIDF,BERT,word2vec,glove. Deeplearning:

Frameworks:

TensorFLow,Pytorch

Neural Network Architecture:

CNN,RNN,Long Short - Term Memory(LSTM)

Computer Vision:

Image Classification,Object Detection,Image Segmentation

NLP:

Sentiment Analysis,text Classification

Language Modeling:

Generative Adversarial Networks

Optimization Techniques:

Gradient Descent,Hyper parameter Tuning

Visualization: Excel,Advanced Excel,Tableau,Matplotlib,Seaborn.

- **GUI Development:** Swing

- **Other Technologies:** Microservices Architecture

**Tools:**

- **Version Control:** Git

Anaconda, Jupyter notebook/Colab, VS code, MYSQL,PostgreSQL,Toad,IntelliJ IDE

-

**Soft Skills:**

- Problem Solving
- Team Collaboration
- Communication
- Time Management

**Recommendation:**
LOR for DATA SCience Internship.

**LANGUAGES**

English,Hindi,Telugu

INNOMATICS RESEARCH LABS

Date: 12th November, 2022

**Letter of Recommendation**

**To whomsoever, it may concern**

With great pride and pleasure, we recommend **Sowjanya Bojja,** who successfully completed his internship in Data Science with Innomatics Research Labs, from July 15, 2022, to October 31, 2022.

Throughout the entire Internship tenure, she has successfully completed the various projects and assignments assigned in time.

Her dedication, determination & zeal to skillfully overcome challenges is striking and are essential traits for being a successful & well-rounded Data Scientist.

**Sowjanya Bojja** is highly deserving and has showcased the willingness to work hard, both efficiently and independently.

We recommend **Sowjanya Bojja** with utmost sincerity and believe that she will be a tremendous asset to your organization.

**Raghu Ram Aduri,**
**Program Manager,**
**Innomatics Research Labs Pvt Ltd**

---

Skills:

- Programming Languages: Python, Java, SQL
- Machine Learning: TensorFlow, Scikit-learn, Keras
- Data Analysis: Pandas, NumPy, Matplotlib, Seaborn
- Natural Language Processing (NLP): NLTK, SpaCy, BERT
- Big Data Tools: Apache Spark, Hadoop
- Web Development: HTML, CSS, JavaScript
- Tools & Technologies: Git, Docker, Jupyter Notebooks

## Some of my Projects:

**Internship projects:**

https://github.com/simplysowj/Internship_july_2022

**Next Word Generator**

- Developed a language model using n-gram techniques to predict the next word in a sentence.
- Utilized the Reuters dataset to build unigrams, bigrams, and trigrams for word prediction.
- Implemented the model in Python and evaluated its performance using accuracy metrics.
- Github link: https://github.com/simplysowj/NLP_projects

**Income Prediction Model**

- Built a machine learning model to predict individual income based on demographic and socioeconomic features.
- Preprocessed and analyzed the data using Pandas, NumPy, and Scikit-learn libraries.
- Achieved [insert accuracy metric] accuracy on the test dataset.
- Github link: https://github.com/simplysowj/Income_pred

**Gold Price Prediction**

- Developed a time series forecasting model to predict the price of gold using historical data.
- Implemented various machine learning algorithms such as ARIMA, LSTM, and Prophet for prediction.
- Evaluated model performance and provided insights to investors for decision-making.
- Linkedin post:https://www.linkedin.com/posts/sowjanya-bojja_internshipjourney-goldpriceprediction-activity-7164817502026321920-712k?utm_source=share&utm_medium=member_desktop
- Github link:https://github.com/simplysowj/Gold-Price-Prediction-including-Flask-app

**Cat vs. Dog Image Classifier**

- Trained a convolutional neural network (CNN) model to classify images of cats and dogs.
- Used TensorFlow and Keras for model development and achieved [insert accuracy metric] accuracy on the validation set.
- Deployed the model using Flask to create a web application for real-time image classification.
- Linkedin post:https://www.linkedin.com/posts/sowjanya-bojja_tensorflow-flask-deeplearning-activity-7163954448375996417-BExp?utm_source=share&utm_medium=member_desktop
- Github link:https://github.com/simplysowj/CatVsDog_Classifier-SpamDetector

**Spam/Ham Email Classifier**

- Developed a binary classification model to distinguish between spam and ham emails.
- Utilized natural language processing (NLP) techniques for text preprocessing and feature extraction.Github:https://github.com/simplysowj/CatVsDog_Classifier-SpamDetector

  **Linkedinpost:**https://www.linkedin.com/posts/sowjanya-bojja_nlp-spamdetection-machinelearng

-activity-7163572648839487490-yET4?utm_source=share&utm_medium=member_desktop

**Microservices for Movie Reviews**

- Implemented microservices architecture to develop a scalable and distributed system for movie reviews.
- Description: Developed a microservices architecture for a movie review platform, enabling scalability and modularity.
- Technologies Used: MongoDB, Spring Boot, React.js, Microservices

https://www.linkedin.com/posts/sowjanya-bojja_i-am-happy-to-share-a-movie-review-application-activity-7160443895137419264-SGYf?utm_source=share&utm_medium=member_desktop

Github link:https://github.com/simplysowj/Microservices

**Salary Prediction Model**

- Developed a regression model to predict salaries based on job attributes and experience levels.
- Utilized linear regression, decision trees, and ensemble methods for model training.
- Conducted feature engineering and cross-validation to optimize model performance.
- https://github.com/simplysowj/Sal_prediction

**Portfolio Management System using LLM**

- Built a personal portfolio management system using the Long-Short Term Memory (LLM) model

https://www.linkedin.com/posts/sowjanya-bojja_i-am-happy-to-share-my-personal-website-activity-7155320719994732544-vHxA?utm_source=share&utm_medium=member_desktop

- https://github.com/simplysowj/Portfolio-sowjanya

**Portfolio Website using React**

- Developed a personal portfolio website using React.js framework.
- Designed responsive and interactive user interfaces for showcasing projects, skills, and experience.
- Implemented routing, state management, and component reuse for efficient development.
- https://github.com/simplysowj/Portfolio

Linkedin post:

- https://www.linkedin.com/posts/sowjanya-bojja_portfolio-webdevelopment-reactjs-activity-7149897737595940864-Lxut?utm_source=share&utm_medium=member_desktop

**Phone Directory Application using React**

- Created a phone directory application using React for managing contact information.
- Implemented features such as add, delete, and search contacts using React hooks and context API.
- Styled the application using CSS and Bootstrap for a modern and intuitive user experience.
- https://github.com/simplysowj/Phone-Directory

**Streamlit-based Project for SQL Queries with Graphs LLM project**

- https://github.com/simplysowj/LLM

**Tableau Project: Bikehaven project**

- Analyzed Bikehaven data using Tableau for insights into users' activity levels, sleep patterns, and health metrics.
- Created interactive dashboards and visualizations to explore trends and correlations in the data.
- Generated actionable insights and recommendations for improving health and fitness goals.
- https://www.linkedin.com/posts/sowjanya-bojja_activity-7146645292652625921-uvu9?utm_source=share&utm_medium=member_desktop
- https://github.com/simplysowj/tableau

**Java Project: Swing GUI Application**

- Designed and implemented a Swing GUI application using Java for a user-friendly desktop interface.
- Developed features such as forms, buttons, menus, and dialogs for data entry and manipulation.
- Applied object-oriented principles and event-driven programming for efficient application development.https://github.com/simplysowj/project-in-java
- Linkedin post:
- https://www.linkedin.com/posts/sowjanya-bojja_i-am-happy-to-share-my-first-java-project-activity-7141257290283683840-gkF1?utm_source=share&utm_medium=member_desktop

**Sql Project :**

IPL Data Analysis using SQL(Postgre sql) and Data visualization using Excel.

https://www.linkedin.com/posts/sowjanya-bojja_successfully-finished-ipl-data-analysis-using-activity-7130554150039257088-LGf1?utm_source=share&utm_medium=member_desktop

**Advanced Excel project:**

https://github.com/simplysowj/fitbit

Analysis of Fitbit fitness tracker data to deliver marketing and business solutions to WeFit and its subsidiaries

https://www.linkedin.com/posts/sowjanya-bojja_analysis-of-fitbit-fitness-tracker-data-to-activity-7116596211960221696-KafI?utm_source=share&utm_medium=member_desktop

**Quora Duplicate Question Detector using BERT and Streamlit**

- Built a duplicate question detection system using BERT (Bidirectional Encoder Representations from Transformers) model.
- Developed a web-based application using Streamlit for users to input questions and receive predictions.
- Fine-tuned pre-trained BERT model on Quora question pairs dataset for identifying duplicate questions.

- https://www.linkedin.com/posts/sowjanya-bojja_bert-streamlit-datasceince-activity-6990799151445807104-ul1C?utm_source=share&utm_medium=member_desktop
- https://www.linkedin.com/posts/sowjanya-bojja_streamlit-heroku-datascience-activity-6989558434228498432--Knu?utm_source=share&utm_medium=member_desktop
- https://github.com/simplysowj/Quora_duplicate_question_detector
- https://github.com/simplysowj/streamlit_quora

**MLflow and Prefect Orchestration**

- Implemented MLflow for managing the machine learning lifecycle, including experiment tracking, model versioning, and deployment.
- Integrated MLflow with Prefect for workflow orchestration and automation of machine learning pipelines.
- Designed and deployed end-to-end machine learning workflows for model training, evaluation, and deployment.
- https://github.com/simplysowj/Prefect_Orchestration_quora
- LinkedinPost:
- https://www.linkedin.com/posts/sowjanya-bojja_prefect-mlops-datascience-activity-6989564940772626432-txdd?utm_source=share&utm_medium=member_desktop
- https://www.linkedin.com/posts/sowjanya-bojja_mlflow-mlops-ml-activity-6989568784457973760-RggX?utm_source=share&utm_medium=member_desktop

**Diamond Price Prediction Application (Heroku + Streamlit +MLOPs)**

- Deployed a diamond price prediction model as a web application on Heroku using Streamlit framework.
- Provided users with a simple interface to input diamond characteristics and receive predicted prices.
- Utilized machine learning algorithms such as regression and ensemble methods for price prediction.https://github.com/simplysowj/Diamond_hiroku
- https://github.com/simplysowj/Diamond_prediction_hiroku

  https://www.linkedin.com/posts/sowjanya-bojja_mlflow-python-ml-activity-6979395069636198400-t3xc?utm_source=share&utm_medium=member_desktop

  a production ready workflow orchestration pipeline using Prefect version 2.4.

  MLOPS task-Diamond Dataset ML pipeline orchestration and experiment tracking.

  The scheduled task runs every one week and has used MLflow to track experiments and for model management.

  https://www.linkedin.com/posts/sowjanya-bojja_mlflow-python-datascience-activity-6977944029040672768-fVrW?utm_source=share&utm_medium=member_desktop

  MLflow is an open source platform for managing the end-to-end machine learning lifecycle. It tackles four primary functions:

Tracking experiments to record and compare parameters and results (MLflow Tracking).

Packaging ML code in a reusable, reproducible form in order to share with other data scientists or transfer to production (MLflow Projects).

Managing and deploying models from a variety of ML libraries to a variety of model serving and inference platforms (MLflow Models).

Providing a central model store to collaboratively manage the full lifecycle of an MLflow Model, including model versioning, stage transitions, and annotations (MLflow Model Registry).

https://www.linkedin.com/posts/sowjanya-bojja_streamlit-machinelearning-dataanalysis-activity-6974940393738035200-JTYx?utm_source=share&utm_medium=member_desktop

**Flask crud app:**

https://www.linkedin.com/posts/sowjanya-bojja_innominion-innomatics-flask-activity-6965630809672204288-kD1N?utm_source=share&utm_medium=member_desktop

**Sentiment Analysis Application (Heroku + Streamlit + BERT)**

- Developed a sentiment analysis web application deployed on Heroku using Streamlit and BERT model.
- Enabled users to input text data and receive sentiment analysis results, including positive, negative, or neutral sentiment.
- Fine-tuned BERT model on sentiment analysis datasets and integrated with Streamlit for real-time inference.
- https://www.linkedin.com/posts/sowjanya-bojja_successfully-finished-covid-19-twitter-activity-6981049587381805056-O1IN?utm_source=share&utm_medium=member_desktop
- https://github.com/simplysowj/sentimental_analysis_Hiroku

**EDA: Ameo 2015 Dataset**

- Conducted Exploratory Data Analysis (EDA) on the Ameo 2015 dataset to understand data distributions, patterns, and relationships.
- Used statistical analysis, data visualization, and descriptive statistics techniques to explore data characteristics.
- Generated insights and recommendations for stakeholders based on EDA findings.

**URL Shortener Application**

- Developed a URL shortener application using Python and Flask framework.
- Implemented URL shortening algorithm and generated short URLs for long input URLs.
- Deployed the application on cloud platforms such as Heroku for public access.
- https://www.linkedin.com/posts/sowjanya-bojja_hi-with-the-help-of-kanav-bansal-i-have-activity-6963560501293301761-mxKc?utm_source=share&utm_medium=member_desktop

**Regex Clone using Flask**

- Created a web-based regular expression (regex) clone using Python and Flask for pattern matching and string manipulation.

- Implemented features such as regex pattern input, string matching, and substitution.
- Deployed the application on a web server for online regex testing and experimentation.
- https://www.linkedin.com/posts/sowjanya-bojja_i-have-successfully-created-clone-of-regex101-activity-6961469088787308544-tzFF?utm_source=share&utm_medium=member_desktop

**Movie Analysis:**

- NLP project using streamlite for sentiment analysis on the
- IMDB Movie reviews tweets dataset, here in this project I used vectorization technique like
- Bert to convert text data into numerical vectors.
- https://www.linkedin.com/posts/sowjanya-bojja_imdb-bert-bert-activity-6981299904170131456-3CAZ?utm_source=share&utm_medium=member_desktop

**COVID-19 Twitter Sentiment Analysis**

- Analyzed Twitter data related to COVID-19 pandemic for sentiment analysis and public opinion mining.
- Classified tweets into positive, negative, or neutral sentiment categories using machine learning techniques.
- Visualized sentiment trends and identified key topics and discussions related to COVID-19 on Twitter.

https://www.linkedin.com/posts/sowjanya-bojja_successfully-finished-covid-19-twitter-activity-6981049587381805056-O1IN?utm_source=share&utm_medium=member_desktop