



Email ID: syamsundarkoya13@gmail.com

Phone No:+91 9959435336

Objective:

A focused and dedicated data science professional with over 3 years of experience in data analysis, machine learning, and natural language processing. Seeking to leverage my skills and expertise to contribute to innovative projects and drive business growth

Professional Summary:

- Over 3+ Years of IT experience in Analysis, Design, Development, and Implementation of various Client/Server.
- Experienced in python data manipulation for loading and extraction as well as with python libraries such as NumPy, and Pandas for data analysis and numerical computations.
- Proficient in machine learning for multiple applications including natural language processing and computer vision.
- Good Knowledge of Supervised and Un-Supervised Learning.
- Good Knowledge on Machine Learning (classification, regression, decision-tree, and Random Forest models).
- Good knowledge of NLP (Text pre-processing and Numerical feature extraction from given text).
- Quick learner with the ability to grasp and learn new technologies.
- Excellent team members having ability to finish tight deadlines and work under pressure. Self-motivated, energetic, and highly ethical in all work-related assignments thus able to immediately contribute to corporate goals and objects.

Working Experience:

- Worked as a **Software Engineer** in Toshiba India Pvt. Ltd., Hyderabad, India from Mar'2021 to Apr'22.
- Working as a **System Engineer** in Tata Consultancy Services (TCS), Hyderabad, India from May'2022 to till date.

Technical Skills:

- **Programming Languages** : Python, and SQL.
- **Statistics** : Various Hypothesis Testing, Estimation, Probability Theory, Time-series Analysis, Statistical Modeling.
- **Machine Learning** : Algorithms for Regression (Linear, Logistic), Classification (Decision Trees, Random Forest, XG Boost, SVM, Naive Bayes, k-NN), PCA, Clustering (K-means, Hierarchical)
- **Natural Language Processing** : Text Processing, Web Scrapping, Sentiment Analysis, Unigram, Bigram and N-gram models.
- **Analytics tools** : Pandas, NumPy, Scikit, Seaborn,Sklearn, Matplotlib, Excel.
- **Cloud Platform** : AWS, GCP, Heroku.
- **Platforms (IDE)** : Visual Studio code, Anaconda, Jupyter Notebook, PyCharm.
- **Source Code Management** : GIT and GitHub, Jira, ALM.

Academic:

- B. Tech (EEE) from SRKR Engineering College, Andhra University, Visakhapatnam with 78.3%.
- Intermediate (M.P.C) from Aditya junior college, Narsapuram with 94.9%
- SSC from Bhashyam English Medium School, Narsapuram with 90%
- Certified for Full Stack Data Science program by Ineuron (Nov'2021-Nov'2022).

Professional Experience:

Project 1:

Project	: Sales Forecasting
Client	: Walmart (UK)
Designation	: Machine Learning Engineer
Duration	: May'22 to till date

Description: Walmart is one of the largest retail companies in the world, and it uses a variety of sales forecasting methods, such as time series analysis, regression analysis, etc., to inform its business decisions. In this article, we will apply regression analysis for sales forecasting on a dataset provided by Walmart.

We will use the daily sales data provided by Walmart, it consists of historical sales data for Walmart stores located in different regions. Each store contains a number of departments, and our objective is to predict the department-wide sales for each store.

Responsibilities:

- My responsibilities to this project are to be involved in Data Exploration, Exploratory Data Analysis (EDA), Data Pre-processing, Data Manipulation, Predictive Modelling.
- Collected, cleaned, and preprocessed historical sales data from multiple CSV files provided by Walmart.
- Managed missing values and handled NULL values in Markdown features by imputing them with zeroes.
- Performed exploratory data analysis (EDA) using visualization tools such as Matplotlib and Seaborn to identify trends and patterns.
- Analyzed the distribution of store types, store sizes, and their impact on weekly sales using box plots and pie charts.
- Investigated the impact of holidays on sales figures through box plots and statistical analysis.
- Evaluated correlations between numerical features and weekly sales using heatmaps to inform feature selection.
- Applied feature scaling using Standard Scaler to standardize numerical features.
- Employed one-hot encoding to transform categorical features into numerical format for model compatibility.
- Split the dataset into training and testing sets with an 80:20 ratio for model evaluation.
- Developed and trained multiple regression models, including KNN Regressor, Decision Tree Regressor, Random Forest Regressor, and XGBoost Regressor.
- Assessed model performance using metrics such as mean absolute error (MAE), root mean squared error (RMSE), and r-squared score.
- Visualized the relationship between observed and predicted sales values using scatter plots to validate model predictions.

Environment: Python, Data Visualization, Descriptive Statistics, Machine Learning, Data Cleaning and Preprocessing, Pandas, NumPy, Matplotlib, Seaborn, Sklearn, Scikit-learn, NLP, AWS.

Project 2:

Project : Future core Employee Prediction
Client : Toshiba
Designation : Machine Learning Engineer
Duration : Mar'21 to April'22.

Description: Toshiba group has four business cores, energy, social infrastructure, electronic devices and digital solutions. Toshiba handled employee performance data carefully, ensuring it was accurate, secure, and private. They used advanced data processing and machine learning to gain valuable insights. By using AI and IoT technologies, Toshiba aimed to improve their HR processes and make informed decisions. This project shows Toshiba's dedication to using cutting-edge technology to boost business success and improve employee performance and know-how they have accumulated in system development and manufacturing. Based on given feature of dataset we need to predict the performance rating of employee.

Responsibilities:

- Collected, cleaned, and pre-processed data from multiple sources to ensure accuracy and completeness for analysis.
- Managed missing values and outliers using statistical techniques such as IQR and Square Root Transformation.
- Conducted univariate, bivariate, and multivariate analysis using Python libraries (Pandas, Matplotlib, Seaborn) to identify key patterns and trends.
- Collected, cleaned, and pre-processed data from multiple sources to ensure accuracy and completeness for analysis.
- Performed hyperparameter tuning to optimize model performance and prevent overfitting.
- Developed and tuned machine learning models using algorithms such as Random Forest, Decision Tree, Gradient Boosting, Linear regression, K-Neighbours classifier, SVM, and Neural Networks.
- Provided actionable insights and recommendations to improve overall employee performance based on model findings.
- Saved the final trained model using pickle for future predictions and real-time application in employee hiring processes.

Environment: Python, Pandas, NumPy, Scikit-learn, Descriptive Statistics, Machine Learning, Matplotlib, Seaborn, Pickle, VS code, Flask, Git, SQL.

Additional Information:

- **Positive Attitude:** To progress positively under all circumstances with conviction and maturity in approach.
- **Adaptability & Flexibility:** To adapt accordingly to the application of adverse or pressure conditions without deviating from critical to quality concerns.
- **Enthusiasm for Learning:** Endlessly strives to learn through different activities.