SIRI SMITHA JOGINAPALLY

1146 Courtney Rd, Apt B, Arbutus, MD-21227 | +1 610-712-0014 | GitHub | LinkedIn
sirismithaj@gmail.com

EDUCATIONAL BACKGROUND

Master's in data science
University of Maryland Baltimore County, MD, US — May 2024
Bachelor of Technology | Information Technology
VNR Vignana Jyothi Institute of Information and Technology, Telangana, India — June 2022

TECHNICAL SKILLS

- Programming Languages : Python, JavaScpirt, C, C++
- Database (DB) : MySQL
- IDE : Visual Studio Code, MySQL, Jupiter Notebook, Google Colab
- Tools : MS Office
- Cloud Platforms : Databricks, AWS, Apache Spark
- Front-End : HTML, CSS, Bootstrap
- Machine Learning Algorithms : Linear Regressing, Logistic Regression, SVM, Random Forest, KNN, K- Means, Naïve Bayes Classifier, Time Series
- Machine Learning Libraries : NumPy, Scipy, Scikt-learn, TensorFlow, Keras, Pandas, Matplotlib, Beautiful Soup, Seaborn

CERTIFICATION

- SQL- MySQL for Data Analytics and Business Analytics, Udemy
- The Data Science Course: Complete Data Science Bootcamp 2023, Udemy
- Apache Spark with Scala – Hands-On with Big Data, Udemy

WORK EXPERIENCE

ACS Solution: Trainee Intern
3 months internship during which I learned MySQL, HTML, CSS, and JavaScript

- SQL – As part of the internship training, we were trained in SQL. We were assigned assignments to implement different SQL queries.
- Web Development – We were also trained in different Web Development Technologies like HTML, CSS, and JavaScript.

ACADEMIC PROJECT

- **Stock Price Prediction using Machine Learning** – In this project, we analyzed the Apple Company Stock Dataset using essential libraries for data manipulation and visualization. After splitting the data, we experimented with various machine learning algorithms, finding LSTM to be the most accurate. Our extended LSTM model predicts Apple's stock price for the next 5 days based on historical data.
  Environment: Google Colab, Python, Visualization, Machine Learning.
  Libraries: Pandas, Numpy, Matplotlib.pyplot, Skearn.preprocessing – MaxMinScalar, Keras.layers – LSTM, Dense, Dropout, Sklearn.model_selection – timeSeriesSplit, Sklearn.metrics – Mean_square_erroe, r2_score, Sklearn – linear_model.

- **Traffic Prediction for Intelligent Transportation Systems Using Machine Learning** – The traffic flow prediction project utilizes date, time, and vehicle count data to forecast conditions for the next 24 hours using recurrent neural network architectures. The most accurate algorithm, evaluated with RMSE and R2 square metrics, will be integrated into a user-interface website, enabling users to input a Zone/Area name and receive a 24-hour traffic forecast.
  Environment: Google Colab, Python, Visualization, Machine Learning.
  Libraries: Pandas, Numpy, sklearn.preprocessing - MinMaxScaler, Mean_squared_error, Matplotlib.pyplot, Tensorflow, Keras, Sequential, Dense, SimpleRNN, LSTM, Dropout, model_from_json, Pickle, GRU, Random

- **Companies Order and Financial Data** – Merging Billing and Customer Details CSV files using Pandas, Seaborn, and Matplotlib, I analyzed a packaging company's dataset. The findings include identifying top box materials, pinpointing the most profitable date, assessing employee performance, highlighting the highest-profit customer, and observing a focus on the national market over the international one.
  Environment: Jupiter Notebook, Python, Visualization, Machine Learning.
  Libraries: Pandas, Seaborn, Matplotlib.pyplot, Numpy, Missingno.

- **Why Delays Increased So Much at EWR in 2022** – Focusing on EWR airport, we analyzed a CSV file using Pandas and Latent Dirichlet Allocation (LDA) to identify key delay factors, including "Traffic management initiatives," "Gate delays," "Taxi delays," and "Airborne delays." The study spanned pre- and post-COVID periods, revealing the pandemic's impact on air traffic and anticipating a return to pre-COVID delay levels due to potential capacity constraints in infrastructure and staffing.
  Environment: Jupiter Notebook, Python, Visualization, Machine Learning.
  Libraries: Pandas, Numpy, Matplotlib.pyplot, Seaborn, Os, Snscrape.modules.twitter, tweepy, re, nltk, stopwords, WordNetLemmatizer, CountVectorizer, LatentDirichletAllocation, TextBlob, WordCloud, LatentDirichletAllocation, CountVectorizer, GridSearchCV from sklearn.feature_extraction.text

- **Sports Apparel Stock Analysis and Forecast** – Analyzing five sports apparel stocks using CAPM, Treynor, and Sharpe ratios, LULU and SKX consistently outperformed, showing positive excess returns and higher Sharpe ratios. Employing SARIMAX for time series analysis, the project forecasts strong performance for LULU and SKX in November, making them attractive investment options in the sports apparel market for investors.
  Environment: Jupiter Notebook, Python, Visualization, Machine Learning.
  Libraries: Pandas, matplotlib.pyplot, numpy, datetime, seaborn, warnings, Fred, sm, train_test_split, yf, mean_absolute_error, mean_squared_error, ARIMA, SARIMAX; sns.set(); yf.pdr_override(); %matplotlib inline; warnings.filterwarnings('ignore').

- **Analyzing and Prediction of Amazon Product Sales Over a Month** – Utilizing 'amazon_categories.csv' and 'amazon_products.csv,' this project analyzed Amazon products through data preprocessing, visualization, and predictive modeling with PySpark. Key insights included highest-selling categories, popular price ranges, and a linear regression model predicting category prices, evaluated by Root Mean Squared Error for accuracy assessment. The study effectively demonstrated data transformation to gain valuable insights into Amazon product trends.
  Environment: Databricks, Python, Visualization, Machine Learning, Apache Spark.
  Libraries: SparkSession from pyspark.sql, functions, desc, col, sum, count, mean, when from pyspark.sql.functions, Window from pyspark.sql.window, VectorAssembler from pyspark.ml.feature, LinearRegression from pyspark.ml.regression, Pipeline from pyspark.ml, RegressionEvaluator from pyspark.ml.evaluation, pandas, matplotlib.pyplot, seaborn, numpy, %matplotlib inline