SHETHALA BOGGARAPU

Jersey City, NJ 07307 | shethalaboggarapu@gmail.com | 551-260-3442 | LinkedIn |Github

Education

Pace University, Seidenberg School of Computer Science and Information Systems	**New York, NY**
Master of Science (MS) in Data Science | **Concentration :** Data Scientist & Data Analyst	December 2023
G Narayanamma Institute of Technology and Sciences(JNTUH)	**Telangana, India**
Bachelor of Computer Science Engineering (Computer Engineering)	July 2021

Summary

Data Scientist with strong expertise in machine learning, time-series econometrics, and financial data analysis. Proficient in Python and SQL for data manipulation and applying advanced techniques like linear algebra and optimization for signal generation and portfolio construction. Skilled in statistics and developing predictive models.

Technical Skills

Programming Languages/scripting Languages: Python, Java, HTML, CSS, XML, PHP, SQL Server, MySQL

Software: VS Code, GIT, Microsoft 365 applications(Excel,Sharepoint, PowerApps, Power Automate), Weka, Jupyter, IBM, Google Cloud Platform(GCP),AWS Sagemaker ,Power BI, Tableau

Distributed Computing: Familiar with Hadoop, Spark, MongoDB, Cassandra, HBase and MapReduces.

Machine Learning Algorithms: Random Forests,k-NN, SVM, Logistic Regression, Decision Forest, Naive Bayes, XGBoost, Fb Prophet,LLMs

Libraries: Pandas, NumPy, Matplotlib, Seaborn, scikit-learn, SciPy, Tensorflow, Pytorch, Keras, OpenCV2,Langchain.

Financial Concepts: Mean-Variance Analysis, Investment Valuation, Asset Pricing Models,Mathematics

Experience

Pearson Vue	Minneapolis,MN
BI(Business Intelligence)/ML Intern	June 23-August 23

- **Data Extraction** and Data Collection is done for 5 years of educational institution data, including course information, using **MicroStrategy**(For security and Privacy Compliance) Databases . Preprocessed the data by analyzing the dataset, Data Cleanup, paying special attention to the impact of covid-19 data.
- Managed outliers through systematic experimentation with different **outlier techniques**, Achieved accurate outlier implementation with **KNN** with **Data Modeling**.
- Collaborated with cross-functional teams to design, develop, and deploy **reinforcement learning** models.
- Leveraged the powerful **XGBoost algorithm**, developed a sophisticated forecasting model with **hyperparameter tuning** and adding lag features enabled to achieve RMSE of 48.56.
- Employed **line graphs** to visually represent the actual versus forecasted values for Infrastructure Management.

Profit Bloom Private Limited	**Telangana,India**
Data Analyst	January 2019-january 2020

- Managed and maintained datasets related to sales and revenue using **SQL** and **Excel**, ensuring data accuracy for portfolio analysis.
- Conducted **advanced analytics** using **TA-Lib** to guide portfolio adjustments based on market trends and volatility indicators.
- Developed end-to-end predictive models with **scikit-learn** and **CatBoost** to forecast profits and revenues, supporting strategic investment decisions.
- Optimized portfolio allocations using **PyPortfolioOpt**, balancing risk and return for enhanced investment performance.
- Built backtesting systems with **Zipline** to validate investment strategies, generating insights for portfolio management.

Globalshala	**Telangana, India**
Data Visualization Intern	August 22- September 22

- Analyzed marketing data for Business Process using Python , understanding business objectives and **visualizing KPI**s to identify underperforming advertising media campaigns for discontinuation.
- **Researching** different **Data Visualization techniques** to create and refine data visualizations in Excel, Tableau, and Python (Jupyter Notebook) based on Process Optimization, enhancing the impact of data presentations.
- Conducted **User acceptance testing** by regular check-in meetings with the project manager to review progress, **refine visualizations**, and ensure alignment with user feedback and project goals.

Smart Bridge	**Telangana, India**
Machine Learning Trainee	June 19 – July 19

- Collaborated in a team of five to **predict global sales** of a video game, using **Jupyter notebook, IBM, & Node Red** and Achieved least mean squared error of 0.5 for Random Forest among **Decision tree, KNN, & Linear Regression.**

Projects

Heart Failure Prediction:
- Analyzed a Kaggle dataset on heart failure using **Python** and **Tableau**, identifying key correlations such as serum creatinine (0.29) and age (0.25).
- Developed a **Tableau** dashboard for **healthcare analytics**, showing age 60 as the highest-risk group, with further insights on platelets and ejection fraction.
- Built and tuned **predictive modeling** for **healthcare** (Logistic Regression, Decision Tree, Random Forest, XGBoost) using **GridSearchCV**, achieving top accuracy with Random Forest (0.89) and XGBoost (0.85) after troubleshooting.
- Applied **ELI5**, **LIME**, and **SHAP** for model interpretability, highlighting serum creatinine and age as significant predictors.

IBM Stock Analysis
- Conducted a detailed **time-series econometric analysis** of IBM stock data (2015-2021), correlating market movements with external economic events.
- Applied **investment valuation techniques** and **mean-variance analysis** to assess stock performance and model **risk-adjusted returns**.
- Developed predictive models using **machine learning algorithms** (KNN), achieving an **85% accuracy** in forecasting stock price movements.
- Leveraged **linear algebra** and **optimization techniques** to refine signal generation and improve model efficiency

MNIST
- Constructed a Sequential model architecture using **Reinforcement learning** comprising a **convolutional layer** followed by **max-pooling**, **flattening**, and **dense layers.**
- Application of **ReLU** activation function for hidden layers and **softmax activation function** for the output layer.
- Implemented custom **stochastic gradient descent (SGD)** optimizer specification with a learning rate of 0.01 and momentum of 0.9 for model training.
- Optimization using **regression testing** facilitated in achieving an **accuracy of 99%** on the MNIST test dataset.

SmartSell
- Conducted comprehensive **root cause analysis** using **dashboards** and **time-series analysis** of sales data from a leading Ecuadorian grocery retailer, including seasonal trends, **user acceptance testing**, and **forecasting** future sales.
- Integrated multiple data sets from different sources, including sales, promotions, and geographic data, to create a unified and **comprehensive analysis** framework that enhanced the accuracy and reliability of the results.
- Analyzed promotional impact using **Tableau Reports** on sales for customer and business requirements, and designed **dashboards** for an Intermediate E-commerce Ecuadorian grocery chain business, revealing findings of an 84% increase in sales during campaigns, emphasizing 'Grocery' as the lead product category, and informing **strategic inventory development** and marketing enhancements.

Time Series Forecasting Using FB Prophet
- Analyzed data sourced from Kaggle regarding number of sun spots on surface over multiple time periods, Developed **Time series forecasting** models in Python to **graphically represent & predict future values**.
- Cleaned data using **Data Manipulation** by removing nulls & changing data types, specified parameters to compare **Statistical Models** such as **Linear & Logistic Regression,** deploying **FB Prophet model** for **exploratory data analysis**.
- Predicted future number of spots with 4.40 of mean absolute percentage error.

Certification

Accenture North America Data Analytics and Visualization Job Simulation on Forage   March 2023
- Merged multiple datasets using **advanced Excel skills/functions** such as **VLOOKUP** to create an integrated and cleaned final dataset, **enhancing data accuracy** and **usability**.
- Developed **data visualizations** and **dynamic dashboards** in **Excel**, facilitating quick interpretation of key performance indicators and aiding in strategic decision-making.
- **Analyzed and identified trends** in top five performing content categories, providing **strategic recommendations** that significantly **improved client engagement** and **operational outcomes.**
- Demonstrated proficiency in **data manipulation** and **visualization tools**, contributing to team capabilities and **business analytics applications** using video presentations for the client and **internal stakeholders.**

Additional Training

Aws Training and Certification- Generative AI(GEN AI) for Executives	June 2024
Aws SageMager- GreatLearning	March 2024
Data Analytics: Dashboards vs. Data Stories(Data Storytelling)	December 2022
PowerBI workshop- Skillslash Academy	July 2022
INSOFE (Data Science Trainee)	Aug 2017 - July 2018