Tharun Kumar Bottlapally

tbottlap@charlotte.edu | +1 (980) 382-7693 | Charlotte, NC | https://www.linkedin.com/in/tharun-kumar-bottlapally/

Summary

Data Scientist with 4+ years of experience in developing and deploying machine learning models to solve complex problems. Proficient in Python and libraries like TensorFlow, PyTorch, and Scikit-learn, with expertise in supervised/unsupervised learning, time series forecasting, and neural networks. Skilled in handling large datasets, big data tools like Hadoop and Spark, and cloud platforms such as AWS, delivering impactful business outcomes through data-driven strategies.

Education

Master of Science, The University of North Carolina at Charlotte   08/2022 - 05/2024 | Charlotte, USA
Data Science and Business Analytics

Technical Skills

- **Languages:** Python, R Studio, SQL, Java, PHP, C, C++, JavaScript, HTML, CSS
- **Database:** MySQL, MongoDB, PostgreSQL, NoSQL, Amazon Redshift, MS SQL Server, Snowflake
- **Data Science:** Machine Learning Algorithms, Deep Learning, NLP, Regression, Classification, Clustering, Recommendation Systems, Neural Networks, TensorFlow, Keras, BERT, ARIMA, Sci-kit Learn, Topic Modeling, FRCNN, YOLO, LSTM, K-Means Clustering, GAN, CNN, RNN, FFNN, NLTK
- **Data Processing & Streaming :** NumPy, Pandas, PySpark/Apache Spark, Hadoop, Apache Kafka, Airflow, Big Data, Big Query, PyTorch, SciPy, SpaCy, Matplotlib, Seaborn, Altair, Plotly, Pycharm, Jupyter Notebook
- **Others & Tools:** Power BI, Tableau, Git, GitHub, Bitbucket, AWS, Microsoft Azure, Django, Streamlit, Flask, Statistics

Professional Experience

Data Scientist, *BatteryXchange*   07/2024 – Present | Charlotte, USA

- Designed and deployed **Machine learning models** using **Python, TensorFlow, Keras,** and **Scikit-learn**, achieving a **25% increase in sales conversions**, a **15% reduction in customer churn**, and a **30% reduction in downtime** by accurately forecasting battery performance and loss patterns, driving operational efficiency.
- Built advanced **Power BI Dashboards** integrated with **cloud-based MySQL** and **AWS RDS**, reducing **reporting times by 40%** and boosting **data accuracy by 30%**, enabling **real-time decision-making** for stakeholders.
- Utilized **NLP tools** like **NLTK** and **SpaCy** to analyze unstructured data, driving customer insights and supporting strategic initiatives.
- **Led full-stack development** and optimized machine learning models for battery performance forecasting, achieving a **30% downtime reduction** and **25% efficiency improvement** using **Docker** and **CI/CD pipelines** to enhance scalability and reliability.
- Proactively managed **large datasets**, implementing **data governance policies** to ensure compliance with **GDPR** and **HIPAA**, maintaining **99.9% data integrity**, and reducing **downtime by 50%** through robust recovery planning and secure backups.

Data Scientist, *Wells Fargo*   01/2024 – 05/2024 | Remote, USA

- Preprocessed and aggregated **10TB of financial data**, resolving quality issues and leveraging **Amazon S3** for scalable storage and **AWS Redshift** for high-performance querying, ensuring data integrity and supporting machine learning models.
- Built **Machine Learning Models** (Random Forest, Gradient Boosting, SVM) and **Deep Learning Models** (LSTM networks, TensorFlow, Keras) for investment predictions and **time-series forecasting**, achieving **92% accuracy** and improving outcomes by **15%**.
- Conducted **feature engineering** using **correlation analysis** and **PCA**, reducing dimensionality and improving predictive accuracy.
- Developed an **interactive dashboard** using **Flask** for real-time investment predictions, reducing analysis time by **40%**, and automated workflows with **Apache Airflow** for seamless data integration and updates.
- Deployed scalable models using **AWS services** like **SageMaker, Lambda, Fargate,and ECS**, with automation via **CloudFormation** and monitoring through **CloudWatch**, ensuring reliability and efficiency.
- **Orchestrated automated ETL pipelines** with **Apache Airflow**, enabling seamless integration and real-time transformation of financial data for model updates.

Data Scientist, *ADP Pvt. Ltd*   10/2021 – 07/2022 | Hyderabad, India
- Built **Machine Learning Models** using **Python**, **TensorFlow**, and **ARIMA** for **workforce planning**, improving staffing accuracy by **20%** and reducing costs by **15%** through **predictive analytics models** and **time-series techniques**.
- Conducted **time-series analysis** with **Pandas** and **Scikit-learn** to identify seasonal trends and forecast resource needs, enhancing **proactive scheduling** and operational efficiency.
- Built ETL pipelines using **Apache Spark**, **PySpark**, and **AWS Redshift**, automating workflows with **Apache Airflow** to process multi-terabyte datasets and reduce data preparation time by **35%**.
- Conducted sentiment analysis on employee feedback using advanced NLP tools (**spaCy, NLTK**), uncovering dissatisfaction trends and delivering actionable insights to HR, resulting in a 10% improvement in engagement scores.
- Developed interactive **Tableau dashboards** to visualize sentiment metrics and workforce performance, enabling real-time tracking and data-driven decision-making by HR teams, while **collaborating with cross-functional teams** to integrate predictive analytics models into enterprise systems for seamless deployment and business alignment.

Data Scientist, *Cognizant Technology*   12/2019 – 08/2021 | Hyderabad, India
- **Developed a data-driven platform** and machine learning ecosystem using **Python, Pandas, NumPy, Scikit-learn, TensorFlow, and Keras**, optimizing business operations, enhancing strategic decision-making, and increasing customer satisfaction by **30%** through predictive analytics.
- Designed **ETL pipelines** to integrate diverse data sources into **AWS Redshift**, ensuring high **data integrity** and enabling efficient access for large-scale analytics.
- Deployed **Machine Learning Models** on **Amazon SageMaker**, automating training, deployment, and monitoring with **CI/CD pipelines**, reducing time-to-market for new models by **40%.**
- Executed **marketing analytics** with **Apache Spark**, identifying customer segments and improving marketing campaigns, leading to a 20% increase in ROI.
- Established a **data governance framework** to ensure compliance with privacy regulations, improving **data quality** and reliability while strengthening system security through collaboration with cross-functional teams.

Projects

JARVIS: The Data Mining Robot
- Developed a virtual robot, JARVIS, to scrape Corporate Action data from multiple websites using Python and automated workflows. Leveraged data mining and processing techniques to analyze and coordinate stock price movements based on extracted Corporate Action data, improving real-time decision-making for end users.

Advanced NLP and Cybersecurity Solutions
- LLM-Based Natural Language Processing: Processed and analyzed over 1 million tweets using advanced NLP techniques. Fine-tuned large language models (LLMs) like BERT and RoBERTa to detect doxing, driving dataset scalability and enhancing privacy-focused AI solutions for improved cybersecurity measures.

Accomplishments

- Published research paper on Job Shifting Prediction and Analysis Using Machine Learning ([Link](#)).
- Smart India Hackathon March 2019 Grand Finalist.
- Presented a business solution at the 2023 Carolina Hurricanes Analytics Challenge.
- Truist 2023 Data Modelling Competition Grand Finalist.
- Published a paper, "Beyond Privacy: Understanding and Mitigating Doxing in the Digital Environment," addressing strategies to combat digital privacy threats.