Sai Vinay Thattukolla

• Tampa, Florida • 813-894-1210 • [saivinay@usf.edu](mailto:saivinay@usf.edu%20) • [LinkedIn](https://www.linkedin.com/in/saivinaythattukolla/) • [GitHub](https://github.com/saivinay111234)

**WORK EXPERIENCE**

# UNIVERSITY OF SOUTH FLORIDA, Tampa, FL Aug 2023 – Present

# Research Assistant

* Engaged in collaborative research projects centered on the development of explanation generation for insider threat, employing generative AI techniques and **hugging face transformer** models for explanations generation in Python.
* Used CERT dataset for insider threat detection, applied LSTM models for event prediction, and employed Large Language Models (LLM) like GPT3.5 and Llama for feature extraction. Utilized SHAP library for feature explanations.

**NYISO,** Albany, NY **May 2023 – Aug 2023**

**DevOps Engineer & IT Infrastructure Intern**

# Developed and deployed a Python application with REST API integration in ServiceNow, automating SGT view addition to CMDB forms & lists, reducing manual workload by 90%.

# Led end-to-end data integration by designing ETL pipelines, consolidating IT data from various sources into a centralized PostgreSQL repository. leveraged Tableau for tracking and reporting, resulting in an 80% reduction in duplicate entries within CMDB tables.

# TATA CONSULTANCY SERVICES, Bangalore, India July 2020 – July 2022

# Data Scientist; Product Management

# Led a team to develop an end-to-end application for customer churn prediction and fraud detection, resulting in a 56% reduction in churn rates and a 41% reduction in fraud cases in Q1.

# Spearheaded the collection of large-scale structured and unstructured data from banking and financial documents, leveraging Tesseract & AWS Sage Maker for efficient data preprocessing and ML pipeline foundation, reducing processing time by 35%.

# Designed and deployed deep learning NLP-OCR model on AWS achieving 89% accuracy in Object Detection using Regex, TensorFlow, and automated it with RPA (Blue Prism).

# Improved NLP model performance to 91% through statistical data analysis, optimizing SQL queries with Amazon Redshift Spectrum to save 23 minutes of daily runtime.

# Created 30+ interactive Power BI dashboards for trends identification and partnered with stakeholders, product managers, UI team, and RPA (Blue Prism) team to develop a banking application integrating ML, AWS, and RPA.

# TATA CONSULTANCY SERVICES, Bangalore, India Jan 2019 – July 2020

# Data Scientist; Supply Chain

# Cleaned, transformed, and analyzed 800,000 consumer records, developing a multi-tiered pricing model through data mining techniques, resulting in a 24% increase in profit margins.

# Conceptualized and implemented a custom ETL process using Oracle SQL, preprocessing the data using Jupyter notebook, handling over 1 million daily transaction records, and storing 4+ petabytes of data.

# Worked with cross-functional stakeholders to prioritize KPIs and analytics needs, creating Tableau reports with ad hoc SQL analysis, leading to $1 million in cost savings. Introduced data-driven solutions, reducing supply chain lead times by 48%.

**EDUCATION**

**University of South Florida Aug 2022 – May 2024**

Master of Business Analytics & Information Systems; Major in Data Science Tampa, FL

**Relevant Courses:** Big Data, Data Science programming, Statistical Data Mining with R, Statistical programming with SAS.

# Velagapudi Ramakrishna Siddhartha Engineering College Aug 2015 – May 2019

Bachelor of Technology; Electronics & Communication Engineering Vijayawada, AP

# SKILLS

# Certifications: Python, SQL for Data Scientist, Machine Learning (Stanford University)

# Tools: RStudio, VS Code, Tableau, Power BI, Qlik, GIT, MS Excel, V-lookups, pivot tables, Jupyter Notebook, MATLAB, Docker, SDLC.

# Languages: C, R, JavaScript, Python (Pandas, NumPy, Scikit-learn, Matplotlib), Scala, SAS, SPSS, Shell Scripting, GIT.

# Databases: MySQL, PostgreSQL, SQL, MongoDB

# Cloud Services: Microsoft Azure, Amazon AWS (S3, Sage Maker, Redshift), ServiceNow.

# Frameworks/OS: Hadoop, Spark, OpenCV, TensorFlow, PyTorch, XGBoost, Windows, Mac, Linux, GCP, A/B Testing.

# ACADEMIC PROJECTS

# Analysis and Prediction of Japan Game Sales on Stock Market (R, RStudio) Jan 2023 – May 2023

# Performed web scraping using Python's Beautiful Soup and R's Rvest libraries to extract the Weekly Nintendo data to analyze the Impact of New Game Releases on Nintendo's Stock Market Returns.

# Employed statistical models including GLM, time series analysis and ARIMA modeling to identify trends and seasonality in Japan sales data, recommended best model with an accuracy of 88.3% by verifying the R-Square value of the models.

**Predict the prices in the stock market by time series regression (Microsoft Azure) Aug 2022** – **Dec 2022**

# Web-scraping the data of the top 100 stock prices in python using yahoo finance, alpha vantage, and pandas and appended the data to the CSV file alphabetically.

# Preprocessed and analyzed stock data using Python; applied logistic regression, decision tree, random forest, LSTM, and multi-layer perceptron models to accurately forecast close prices, achieving a 71% accuracy for the LSTM model.