Santhiya Theanraj

santhiya.theanraj@gmail.com 609-332-8098 Santhiya Theanraj santhean

Education

Temple University - Philadelphia,US *Aug 2024 – May 2026*
MS in computational data science

Anna University - Chennai,India *Aug 2011 – May 2015*
Bachelors of technology in information technology

Indian Institute of Management Bangalore - Bangalore,India *Jun 2018 – Jun 2019*
Post graduate certificate program in business intelligence and analytics

Technologies

Tools: Incorta, SAP BODS, SAP BO, PowerBI, Informatica Power Center, Cognos Data Manager, Cognos BO Reporting, Jupyter Notebook, Git,Oracle, MS SQL Server, AS400, Teradata, MongoDb, Hadoop File system
Languages: Python 3.4,SQL, Pyspark,Pytorch,SQL, Bash scripting

Experience

Advanced software Engineer *Chennai, India*
Light and Wonder *Nov 2018 – Aug 2024*

- Managed and maintained multiple ETL tools, including SAP BODS, Cognos Data Manager (legacy), and SSIS, ensuring smooth data integration across systems.
- Took ownership of Cognos Data Manager ETL processes, overseeing issue resolution, new development, and maintenance for production pipelines, including weekend and month-end support.
- Automated month-end data loads using batch scripts, reducing manual intervention and improving reliability.
- Developed SAP BODS jobs, including an SFTP-based ETL pipeline to extract Excel-based lottery files from third-party systems and load them into internal systems.
- Designed and implemented multiple ETL workflows in SAP BODS for system migration projects.
- Migrated Cognos BO reports to SAP BO, including complex solutions such as the *Deal Approval Form*—integrating Live Office in Excel with WebI reports and dynamic updates via Power Query.
- Optimized query performance and improved ETL runtime by **40%**, significantly enhancing processing efficiency.
- Developed Incorta reports for Accounts Receivable (AR) and Accounts Payable (AP) analysis.
- Built Power BI dashboards for team performance tracking, including ticket status monitoring and workload distribution.

Senior Software Engineer *Chennai, India*
Wipro Limited *May 2015 – Nov 2018*

- CVS Caremark-Developed ETL pipelines using Informatica powerCenter to transform and move data between systems. This likely included mapping out data flows, defining transformations, and orchestrating the movement of data. Used Teradata as datawarehouse platform, developed shell scripts to automate the ETL process.
- Cargill Foods-Predominantly worked on modeling business data pipelines into a data warehouse for Cargill EMEA regions. Worked with a variety of database systems, including AS400 and Oracle, coordinated with cross-functional teams to ensure data alignment between different departments. Used Informatica Powercenter as key ETL tool and worked as a primary data engineer in version migration and directly coordinated with business.

Projects

Multi-labeling classifications of documents(IIMB)

- Provided with pickle data set which contains mails and documents in table form in deutsch language with class imbalance issues
- Used Word and sentence tokenization, stemming and lemmatization for data preprocessing.Addressed class imbalance by grouping target labels and SMOTE method.Used algorithms like Naïve Bayes, SVM Model, Logistic Regressions, LSTM, RNN.
- Tools used: Python 3.4, Jupyter Notebook, Anaconda.

Comparative analysis of hybrid CNN-SVM machine learning Vs transfer learning for crop pest classification(Temple)

- In this project a comparative study on how efficient SVM works for small image dataset over pretrained models like VGG16, Resnet50, Efficient Net models
- Used different optimizers like SGD, ADAM, ADOPT to study the efficiency of each optimizer and ADAM outperforms other optimizers
- Challenges in this project is the dataset which is sourced from web where we had 25 classes different pests with class imbalance. Used Pytorch Image Augmentation to up sample the data which really helps in improving the overall accuracy
- Tools Used: Python, Jupyter notebook, Google collab, Kaggle notebook

Discovering Unknown Side Effects of Multi-Drug Combination Therapy using Link Prediction(Temple)

- Designed a link prediction framework for drug-drug interactions (DDIs) to identify potential unknown side effects of multi-drug therapies.
- Implemented traditional network-based approaches alongside modern machine learning techniques, including Node2Vec fusion models, Graph Convolutional Networks (GCNs), and Multi-Semantic Transmission Embedding (MSTE) with heterogeneous graphs.
- Achieved **80% prediction accuracy** using the MSTE model, demonstrating the potential of graph-based methods in pharmacovigilance.
- Tools Used: Python, PyTorch Geometric, NetworkX, Node2Vec, Graph Convolutional Networks (GCNs), Knowledge Graph Embeddings (MSTE)