Charan Teja D

9740320024 | dvrscharanteja@gmail.com | Bengaluru, karnataka

Summary

Senior Data Scientist familiar with gathering, cleaning and organizing data for building machine learning pipelines. Highly motivated to work in the fields of analytics and data science that can enhance the understanding of data and extract meaningful insights. Eager to bring 8.5 years of experience to solve interesting business problems.

Skills

- **Machine Learning** – Regression, Classification, Time series, Deep Learning, NLP, LLMs
- **Use cases** – Forecast, Customer segmentation, Price Elasticity, Churn Prediction, Market Basket Analysis, Inventory management.
- **Languages** – Python, R, SQL, PySpark
- **Databases** – PostgreSQL, SQLserver, OracleDB, Hadoop, DynamoDB, pgvector
- **AWS** – EC2, Lambda, Sagemaker, Kendra, Textraxt, Bedrock, IAM roles
- **Tools** – Alteryx, Pentaho, Docker, Github
- **Frameworks** – Anaconda, Django
- **MLOps** – Airflow, MLflow
- **Python libraries** – pandas, numpy, tensorflow, keras, OpenCV, tesseract, easyocr, sklearn, statsmodel
- **LLMs** – Cohere, Titan, Claude, Llama
- **BI tool** – PowerBI

Education

Indian Institute of Technology madras | Chennai, Tamilnadu
Aerospace Engineering | 07/2013

- Major - Aerospace, Minor - Mathematics

Experience

Prescience Decision Solutions | Bengaluru, Karnataka
Senior Data Scientist | 03/2023 - Present

- **Machine Learning Framework for Retail use cases**
  - Led a team of 3 to build ML models for retail use cases such as **forecasting, customer segmentation, Inventory management, Market Basket Analysis, Churn prediction.**
  - Streamlined Quality checks and EDA processes to reduce turnaround time for preprocessing by **40%.**
  - Utilized **Airflow** to deploy Models, Monitor performance, Schedule runs and track errors that reduced ML lifecycle turnaround time by **30%**.
  - **Technologies** - scikit-learn, prophet, statsmodels, PostgreSQL, Mysql, Airflow, pandas, pyspark, Docker, PowerBI.
- **Revenue forecasting by predicting for opportunity conversion.**
  - Led a team of 2 to develop pipeline for predicting the outcome of an opportunity and estimating revenue based on it.
  - ROC and AUC analysis and grid search are used to arrive at best performing model.
  - Boosted prediction accuracy to **93%** that helped track performance of teams and plan for each quarter.
  - **Technologies** - python, scikit-learn, SMOTE technique, XGboost classifier.
- **LLM based Workflow for querying on top of existing knowledge base**
  - Led a team of 2 to build LLM based solution to obtain key metrics from vendor agreements
  - Workflow eliminated need for manual resources in evaluating performance of vendors and monitor changes in time.
  - Lowered the turnaround time by over **80%** by automating the process.
  - Leveraged tools such as Amazon Titan Embeddings, Postgre, **Retrieval-augmented generation (RAG)** techniques and LLMs present on AWS Bedrock to extract the key metrics.
  - **Technologies** - AWS Bedrock, AWS Textract, LLMs (Cohere, Llama, Titan), Amazon Sagemaker
- **Develop workflow to extract, forecast, automate and validate the tax liabilities**
  - Led a team of 2 to extract data from tax documents using **OCR**, **Multivariate time series** to estimate for tax liabilites and **Alteryx** workflow to streamline and validate the calculations.
  - Streamlined the process by eliminating need for manual effort that reduced turnaround time by **95%**

- **Technologies** - Python, OpenCV, tesseract, easyocr, Alteryx, Docker, Time series.

Prescience Decision Solutions | Bengaluru, Karnataka
Data Scientist | 03/2021 - 02/2023

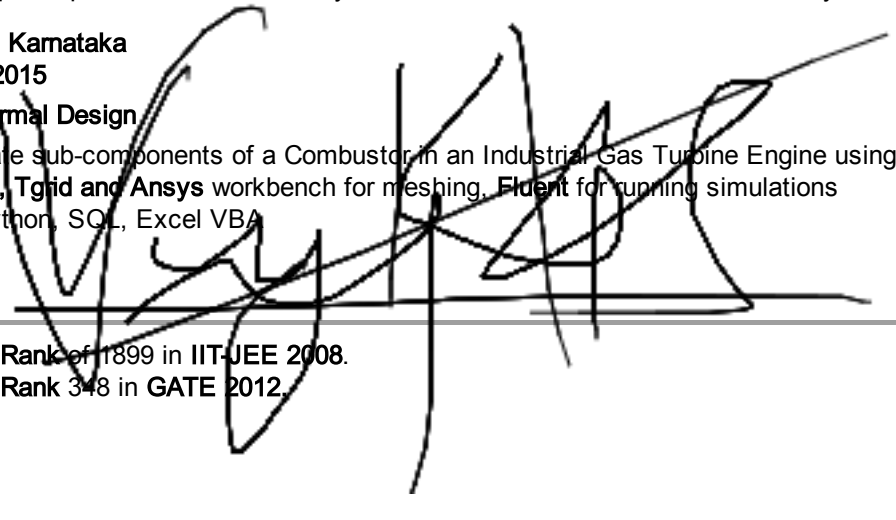
- **Cross sell recommendations model for targeted e-mail campaigns (e-commerce)**
  - Use various **collaborative filtering** models such as **cosine similarity**, **matrix factorization** along with **Deep Auto Encoders** for cross sell product recommendations, which helped to increase revenue by **13%**.
  - Feature Building on top of data from **oracle** using **SQL** and **python**.
  - **A/B testing** methodology used, to find effectiveness of models used in email campaign
  - **Hyperparameter tuning**, **champion- challenger** framework to improve accuracies in next campaign cycle.
- **Demand Forecasting & Inventory Management (e-commerce)**
  - Use **multivariate time series** model to forecast number of orders that is to be used for capacity utilization planning, SKU level predictions and stock quantity recommendations.
  - **Extract Transform and Load (ETL)** data from multiple sources like **PostgreSQL** and **Redshift**
  - Perform **Exploratory Data Analysis (EDA)** which include imputation, Principal Component Analysis (PCA) and outlier detection and removal, to converge on features that can be used for model building.
  - Manage ML lifecycle by deploying models, tracking performance, schedule runs and evaluate outputs using **MLflow**
- **Promo effectiveness & Market Mix model (e-commerce)**
  - Mapping transactions with promos using data extracted from **Redshift**
  - Evaluated Performance of various promo channels used currently to classify them based on revenue generation.
  - Use **Linear regression model** to arrive at optimal spending on various marketing channels to improve margins by **28%**
- **Customer propensity to pay score (Fintech)**
  - Used **XGBoost algorithm** to score customers and streamline enquiry process of customer support that improved loan recoveries by **20%**.
  - Feature Engineering to clean data, remove outliers, impute missing values and transform tables in **Pentaho** on top of data extracted from **postgreSQL**
  - **Update deployed model** with new data every week and monitor the performance of models based on recovery rates.
- **Technologies -** Pyspark, Tensorflow, Keras, scikit-learn, Statsmodels, MLflow, OracleDB, Redshift, PostgreSQL, Pandas

Prescience Decision Solutions | Bengaluru, Karnataka
Senior Data Analyst | 06/2018 - 02/2021

- **Channel Partner Analytics (Manufacturing firm)**
  - Developed **Machine Learning** models using **Logistic Regression & Decision Tree**, to rate current channel partners and improve channel partner onboarding process.
  - Use of **MICE algorithm** for data imputation and **PCA** to remove multicollinearity in input dataset.
  - Worked with stakeholders to develop quarterly roadmaps based on impact, effort and KPIs
  - **Technologies** - Pandas, scikit-learn, Excel, python, MySQL
- **Image Processing to extract and analyze data**
  - Developed **OCR tool** using **OpenCV, pypdf and easyocr** for tax team in e-commerce firm to extract relevant tables from pdfs and images that are stored in **S3**, which reduced manual effort and improved turn around time by **90%.**
  - Use of Image processing modules such as **Tesseract** and **AWS Textract** to develop an **analytics pipeline** for a Banking firm that reduced turnaround time by **80%**.
  - Developed process to automate relevant data extraction from unstructured legal documents to be stored in **MySQL.**
  - **Classification** of e-mails by processing attachments using **OCR** and performing analysis on extracted data using **regex,** that reduced sorting time by 90%.
  - **Technologies** - pypdf, OpenCV, Tesseract, easyocr, S3, AWS Lambda, AWS Textract, MySQL, elasticsearch.

GE Aviation | Bengaluru, Karnataka
Engineer | 07/2013 - 09/2015

- **Combustion AeroThermal Design**
  - **Design** and Validate sub-components of a Combustor in an Industrial Gas Turbine Engine using **Siemens NX** for modeling, **Gambit, Tgrid and Ansys** workbench for meshing, **Fluent** for running simulations
  - **Technologies** - Python, SQL, Excel VBA



Achievements

- Secured **All India Rank** of 1899 in **IIT-JEE 2008**.
- Secured **All India Rank** 348 in **GATE 2012**.