**DATA SCIENCE PROJECTS**

**Deep Learning - Image Recognition**

Application of Transfer Learning in various use cases. Currently working on detecting diseases in an XRay by using Resnet architecture and fine tuning with provided dataset. Productionizing models on AWS Lambda platform and using Python starlette package. Have successfully applied transfer learning on Geospatial images and finding existence of various objects.

Technical Environment: Pytorch, Fastai, AWS

**Vibration Data Analysis and Condition Monitoring**

**Merck, Inc.**

Condition monitoring by analysis of Vibration data for determining the health of a device. Time waveform, Spectrum Analysis, Fourier Transform are some of the techniques used here. Designing rules based on statistical considerations such as movement of orders, persona changes, etc. Setting up the architecture for scaling of Fault Detection, interpretation of faults, etc. Further suggested and developed framework for predictive models to determine the probability of failure of a component.

Technical Environment: RStudio, dplyr, ggplot2, plotly, postgres

**Natural Language Processing – Financial Modeling**

Prediction of alpha based on sentiment and stock price movement data. Processing of financial articles and determining sentiments attached to it, using NLTK library (sentient and vader). External sentiment data was correlated with Stock movement data and subsequently, machine learning models were built. These models were made accessible using Flask in Python. Also a search mechanism was developed using TF-IDF models and cosine similarity scores. Given a stock and sentiment polarity, most relevant articles related to stocks are found. Models created were operationalized with help of the Flask library available in python.

Technical Environment: Python, NLTK, Flask

**Cummins Inc - Data Science**

Cummins Inc has worked on data science initiatives to bring down costs related to repair and maintenance. I have set up the data pipeline which reads data from various tables and creates a single dataset which helps in building machine learning models to determine probability of component failure given a symptom and service model name. This helps in maintaining appropriate stock at various warehouses.

Second use case is to find the reasons for repeated visits. Repeated visits are a source of revenue impact as warranty is breached when an engine comes for repair within a month of earlier visit. This is a data discovery exercise where I am writing R Script to generate a dataset which contains reason, symptoms related to a repeated visit and find probable causes.

**Internet of Things**

Worked on a number of RFPs related to Data Science proposals. Proposed the usage of Weibull distribution and others to estimate the life of complex equipment like Power Transformers.

Worked as a data scientist on an anomaly detection offering for various use cases. Worked on Chiller Plant (HVAC) data to identify anomalies.

Worked on a pilot for Manufacturing client to gather insights from data using exploratory data analysis. This was targeted to identify production lines with good OEE (Overall Equipment Efficiency) and ways to maximize it.

Designed a predictive maintenance solution around Aircraft maintenance. Data Science related use case was defined, designed and executed by me using the R Statistical Package. This demo has been presented in International Airshow at Farnborough, UK.

Designed an IOT Solution for connected devices. This project consists of developing a generic simulator and designing the interface for connected products. Implemented predictive analytics by applying machine learning on machine and asset related data, batch alerts using R. The connected products solution has been developed in a way that it can cater to a number of domains and clients. Have participated in creating demos for a number of clients.

Implemented Predictive maintenance for Connected Products using R Programming language for condition monitoring. Combined device and asset data to create a labeled dataset which was later used to determine probability of failure of a component in the next fifteen days.

Technical Environment: R Programming Language, Python, IBM Bluemix, IOT Foundation, Node JS, Azure MLStudio.

**Building Big Data Analytic Pipeline**

Worked as Senior Technical Architect on designing an Analytics pipeline for real time streaming with device data.

Technologies used are R Programming Language, Spark Streaming, Spark Machine Learning libraries. Also, building various models, such as Decision Tree Model, Random Forest model for implementing various use cases related to IoT.

Technical Environment: Scala, Spark, Hadoop and Kafka.

# Telecom Analytics

Worked as a Data Scientist with Big Data Analytics. This profile required planning and execution of various proof of concepts related to analytics, such as fraud detection, roaming analytics, etc.

Have worked on a PoC for identifying fraud related to Simbox. This required feature generation from existing call detail records and then identifying fraudulent machine and sim ids.

Have worked on CRM records and Voice and SMS data for identification of silent roamers and provide recommendations for increase in revenue.

Worked on a PoC for reducing the revenue loss for premium numbers. This PoC requires in correctly identifying the users likely to default.

Tools - Hadoop, Mapreduce, Tableau and R (including ggplot).

**I-Analytics**

This project was undertaken to build various statistical models using the concepts of statistics and machine learning (Random forests, K-Means Clustering, Logistic regression).

These models have been developed to address a variety of use cases including:

Designing a sentiment mining system by applying Naïve Bayes principles.

Development of use cases and Analysis of crime related data.

Development of a multivariate predictive model for doing spend analysis.

Development of use cases for using Hadoop as an archival tool.

Role: Designing Data Science related implementation, create machine learning models.

Responsibilities:

Analysis of data using standard statistical models (linear regression) and machine learning concepts (K means, Random Forests, etc.)

Planning and executing POCs.

Resource mentoring on Big Data and Analytics.

Tools : R, Hadoop, Hive, Pig.

**Project – I-Port**

This project has been initiated to develop a product I-Port whose aim is to create a dashboard and incorporate forecasting models leveraging Big Data and advanced analytics.This product has the capability to provide a single view of the organizational processes and their performances., A number of KPIs have been developed using various technologies.

Creation of BI dashboard for BPO processes for various KPIs. These include Fraud chargebacks, Call volume, calls trend, etc. Development of predictive models for calls and complaints. Qlikview design and development of complex dashboards and reporting.

Developing forecasting models (uni and multivariate) using R. Implementation of forecasts on a dashboard.

Designing a model for doing sentiment analysis (location specific) using twitter.

Developing the architectural solution involving Hadoop (hive) and Qlikview.

Role: Data Analyst/ Consultant.

Responsibilities:

Developing the technical architecture of the product. Evaluating Big Data and open source technologies and building the proof of concept.

Development of forecasting model for predicting volume of calls (using Pig and R).

Writing PIG/Mapreduce programs for summarizing the data.

Designing a model of sentiment analysis using twitter data.

Achievements:

This product (developed in house) has addressed the need of having a dashboard with predictive capabilities. This solution has used Big Data technologies thus building organizational capability.

Developing a BI dashboard using Qlikview with no prior capability. This is a fully deployed solution in Qlikview using SAP extracts and Qlikview 11.

Environment: MySQL, Talend, Pentaho, Hadoop, Pig, R (Regression), Qlikview.

**EMPLOYMENT SUMMARY**

Pentair Water India Ltd **Feb 2022 - Till Date**

HCL Technologies Limited **May 2015 - Feb 2022**

Mobileum India P. Ltd. **Sep 2014 - Jan 2015**

Steria India Ltd. **Aug 2004 - Sep 2014**

Perot Systems **Feb 2003 - Jul 2004**

TCG Software Services Pvt Ltd **Jul 1998 - Feb 2003**