

PRADYUMNA DOMMATA

pradyum2104@gmail.com

517-303-7947

Summary

Data Science professional with around 5+ years of experience in Supply Chain, Marketing, Automotive and IT domains, following CRISP-DM methodology which involved gathering requirements by working with cross functional stakeholders, performing Data Extraction, Data Screening, Data Wrangling, Data Exploration and Data Visualization of structured and unstructured datasets as well as implementing extensive Machine Learning & Time Series algorithms to deliver resourceful methods, insights, inferences that significantly impacted business revenues and user experience's.

Skills

- | | | |
|---------------------|----------------------------------|--------------------------------|
| • Python | • PySpark, Hadoop | • MATLAB, Time Series |
| • SQL, NoSQL | • AWS - Sagemaker Athena, Lambda | • MS Excel, Access, Visio |
| • Azure – ML Studio | • IBM Aspera Cloud | • Statistics, Machine Learning |
| • Tableau, Power BI | • PLC, GE Proficy, MES | • Git, Flask API |

Experience

Microsoft

Data Scientist (Contract)

Chicago, IL

March 2021 - present

Responsibilities:

- As a part of Field Marketing & Operations Team, was involved in **Resource Utilization Forecasting** and **Strategic Capacity Planning Optimization** projects.
- Extracted the required data from various **Microsoft SQL server** databases for data munging in python notebooks with the help of **pyodbc** connectors and exported the transformed data to **Azure SQL Database**.
- Analyzed the transformed datasets and addressed the anomalies in **Azure ML Studio** python notebooks.
- Forecasted resource utilization using ensemble of ML algorithms like Auto-ARIMA, **FB Prophet** and shortlisted the best forecast algorithm for each categorical cross-section in the data using model & business KPI's.
- Explored the resource capacity planning business use case, performed **literature survey**, participated in **brain storming sessions** along with the team and came up with different **optimization** models.
- Analyzed the capacity planning related datasets in **Power BI** and python notebooks, **handled anomalies** using an **algorithmic approach** and tested different planning models formulated by the team.

Environments: Python, SQL, Microsoft SQL Server, Azure ML Studio, Power BI, Excel.

Koch Industries

Data Scientist (Contract)

Chicago, IL

July 2019 – February 2021

Responsibilities:

- Gathered **business requirements**, converted them to technical requirements and worked with **cross functional stakeholders** to understand existing supply chain process flows and identified the key zones to be tracked for driving business outcomes. Ran the stakeholder meetings in the absence of Project Manager.
- Performed **Business process modelling** (BPMN) in **MS Visio** in order to analyze and automate the existing practices.
- Primarily worked on projects which involved improving the **demand forecasting** accuracy of the SKU's by following CRISP-DM methodology and using Advanced Time Series ML models.
- Have setup a data ingestion pipeline for SKU level Data Extraction from AWS **S3** and AWS **Athena** using SQL queries. Explored and manipulated large **RDD Datasets** from legacy Hadoop cluster using **PySpark**.
- Munged the data to several formats in order to conduct large scale data analysis and modeling to derive actionable insights and provided recommendations on metrics to be tracked.
- Performed EDA on time series data in **AWS Sagemaker** notebooks to find underlying data issues, handled imbalanced datasets by **resampling** and **interpolation** and checked for **descriptive** and **prescriptive** statistics.
- Performed **Feature Engineering** on internal data features and **Feature selection** on numerous Macroeconomic external features by ranking them using algorithms like **XGBoost**.
- Performed **Dimensionality Reduction** using **Principal Component Analysis** and other methods.

- Performed **classification** of time series data points based on certain conditions like **SKU life cycle**.
- Clustered time series data points using **K-Means clustering** into groups based on conditions like **SKU life cycle, Customer Segmentation, velocity, volatility** and performed analysis on cluster groups to choose the best model.
- Assisted Senior Data Scientists in applying statistical time series and machine learning models like **ARIMA**, RNN based deep learning model like **DeepAR** on the transformed datasets to produce forecasts.
- Performed **batch processing HPO jobs** within a lag X region and analyzed the results to select the stable models for accuracy measurement.
- Performed **A/B Testing** between the predictive models within a lag across 4 business regions to come up with a stable and most accurate model to implement in production.
- Designed and developed KPI's like **Bias, Accuracy, Dollarized Weight, Value Add** at various levels and reports around supply chain performance, dashboards, and analysis for model selection in order to drive key business decisions.
- Calculated error metrics like **MAD, WMAPE** at various levels for the models based on generated forecasts.
- Worked with team to conduct **root cause analysis**, compared the metrics with baseline models, reported accuracy differences and communicated the best models to implement.
- Used **Tableau** and **AWS QuickSight** to present insights to stakeholders in the form of user stories and dashboards.
- Evaluated forecast quality, model KPI's, measurement metrics and finally integrated the forecasts with **SAP APO** and helped to configure the entire process flow with **AWS step functions, AWS Lambda, AWS Glue** and **Gitlab CI/CD**.
- Reported an accuracy boost of **6.3%** in my recent engagement which resulted in **\$12M** annual revenue boost for the organization.

Environments: Python, SQL, PySpark, Tableau, AWS - Sagemaker, S3, Athena, Lambda, QuickSight, Gitlab, Excel, Visio.

General Motors
Data Science Analyst

Lansing, MI
August 2017 - May 2019

Responsibilities:

- Gathered business requirements, converted them to technical requirements and communicated with stakeholders for better understanding of the process flows and areas of improvement.
- Have setup a data ingestion pipeline to log and store real time drive data from vehicle's Electronic Control Unit (ECU) involving **On Board Diagnostics (OBD)** Data and the **CAN (Controller Area Network)** Data in the **IBM Aspera Cloud** Database in **Measurable Data Format (MDF, MF4)**.
- Munged the data into csv format to perform **Exploratory Data Analysis (EDA)** on CAN Data in **MATLAB** involving **Arbitration, Re-Sampling of CAN Signals, Stochastic Signal Processing** and **Sorting channels** based on Bus Sources & Frequencies.
- Calculated **channel metrics** and **percentage statistics** of the channels present in MF4 files after initially converting the data into **Tall Arrays**.
- Formatted Tall Array data in order to act as an input to the integration models or subsets of it in the form of **UML** diagrams in **IBM Rational Rhapsody** for playback open loop testing.
- Developed a **MATLAB API** for the client in order to deploy it in their cloud server. The API was able to classify and label existing data and map it to generic structure and metadata of **System Behavioral Tests**. The API was also able to perform EDA based on historic data, report trends and produced forecasts.
- Have used the Time Series Models like **ARIMA**, Deep Learning models like LSTM's using TensorFlow and keras in the backend of the API to produce forecasts.
- In the second phase, I have scrapped out the **Power Train Control Module (PCM)** Data from Aspera Cloud using **Aspera Managed File Transfer (Aspera Connect)** and performed analytics in **PyCharm** using python in order to find the relationship between the **Diagnostic Trouble Codes (DTC)** and Sensor Data.
- Performed EDA on the data including univariate, bi-variate analysis using **plotly** for each trouble code and found the correlations between the sensor data using **Seaborn** and predicted the future status of the sensor signals by applying the time series models like **SARIMA** and **GARCH**.
- Found the sensor signals responsible for the Trouble Codes using EDA, Correlation Analysis & Predictions.
- The results were summarized as a dashboard in **Tableau** and are presented to the stakeholders.

Environments: IBM Aspera Cloud, MDF, MF4, MATLAB, IBM Aspera Connect, PyCharm, Python, Tableau, PowerBI.

Responsibilities:

- Responsible for requirement gathering from the client and presenting them in the stakeholder meetings.
- Remotely configured the production lines and manufacturing machines of P&G using **GE-Proficy** which is an **Industry 4.0** MES software.
- Implemented solutions like Line Event Detection System (**LEDS**) and Quality Analysis (**QA**) on the production lines using GE-Proficy.
- Performed validation of production lines and manufacturing machines using IQ, OQ and PQ documents.
- Configured the **PLC** ports using **Proficy Historian** tags for data collection from various sensors embedded in manufacturing machines and production lines.
- Worked with **VLOOKUP's, macros, pivot tables, formulas, and conditional formatting** to summarize data in Excel.
- Performed SQL queries on Proficy Database known as Proficy Historian to extract and analyze the manufacturing process data and wrote stored procedures to implement process flows.
- Performed Exploratory Data Analysis (EDA) using **Tableau** on the SKU and Plant Level Time Series Data and provided insights which reduced downtime and improved efficiency of the Manufacturing Plants.
- Configured Views in **Proficy Plant Applications (PPA)** to report **KPI's**, depict real time process flows of manufacturing process. Reported key insights in the form of user stories and dashboards using **Tableau**.
- Trained fellow associates on **Manufacturing Operations Management** and GE-Proficy.

Environments: SQL, Tableau, GE Proficy, Proficy Plant Applications, Advanced MS Excel.

AMMACHI Labs
Intern / Data Analyst

Kerala, India
December 2015 - May 2016

Responsibilities:

- Developed a **PLC** code to control the process flow in **Automation Studio, RS Logix 500**, performed integration and testing of logic on an **Allen Bradley PLC**.
- Interfaced PLC controller with **load cells**, safety cut off relays and calibrated them for efficient operation.
- Performed trials to study the effectiveness of pneumatic grippers on **IRB 1600 ABB** robot test bench.
- Developed the end-to-end simulation model in **MATLAB Simulink** and published a research paper on it as a first author in ICAARS - 2016.
- Extracted PLC data using SQL queries and performed **EDA** (Exploratory Data Analysis) on it for sensor calibration, testing and explore ways to improve the efficiency of the process flow.
- Analyzed the data in **Excel** using **VLOOKUP's, macros** and summarized it using **pivot tables** and charts.
- Presented results to the stakeholders in the form of graphs and presentation decks.

Environments: SQL, MATLAB, Automation Studio, MS Excel, MS PowerPoint.

Education

Michigan State University
Master of Science, Electrical Engineering

East Lansing, MI
August 2017 – May 2019

Amrita University
Bachelor of Technology, Electrical & Electronics Engineering

Kollam, India
August 2012 – May 2016

WEB Links

- **LinkedIn** – <https://www.linkedin.com/in/pradyumna-reddy-36a658ab/>
- **GitHub** - <https://github.com/Pradyum2104>
- **Kaggle** - <https://www.kaggle.com/dommatap>