**PRADYUMNA REDDY**Data Science Consultant

**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.

* Experienced in facilitating the lifecycle of a Data Project: gathering **Business Requirements**, communicating with **cross functional stakeholders, Business Process Modelling, Data Extraction, Data Pre-Processing, Feature Engineering, Dimensionality Reduction, Algorithm Implementation, Back Testing** and **Validation**.
* Experienced in working with statistical hypothesis tests involving parametric tests like **t-test, z-test, ANOVA tests** along with non-parametric tests like **chi-sq. tests, Mann-Whitney U** and **Wilcoxon rank tests.**
* Adept at analysis of missing data by coming up with unique **Statistical Imputation** methods, exploring **correlations** and similarities, introducing **dummy variables** for missingness, and choosing from imputation methods such as **mean, median, mode Imputation** & multivariate Imputation methods like **Iterative Imputer** in Python.
* In-depth Knowledge of **Dimensionality Reduction** methodslike **PCA, LDA** and **model regularization methods** like Lasso & Ridge Regression.
* Proficient in data collection from various sources & performing **Exploratory Data Analysis**, quantitative and qualitative data analysis, aggregating the data, finding **patterns, outliers**, and key insights in the data sets.
* Adept in performing **A/B Testing** between predictive models or cluster groups by laying out hypothesis, filters on the datasets, defining success metrics, adjusting sample size & checking for statistical significance.
* Experienced in supervised ML Techniques such as Regression and Classification Models like **Linear, Polynomial, Logistic Regression, Decision Tree Classifier, Support Vector Machines** and **k-NN** algorithms.
* Experienced in ensemble learning methods like **Bagging, Boosting** and unsupervised **Clustering** algorithms like **K-means, C-means,** and **Association Rule Mining** with **A-priori Algorithm.**
* Experienced in Time Series Forecasting methods like **ARIMA, RNN** based **DeepAR** and **Stochastic** methods.
* Experienced with Python libraries like **NumPy, pandas, datetime, tslearn, statsmodels, sklearn,** and **SciPy**.
* Experienced in model optimization using hyper parameter tuning methods like **Grid Search CV.**
* Skilled in Big Data Technologies like **Microsoft SQL Server, MySQL, Spark SQL, PySpark** and **Databricks.**
* Adept with Cloud Technologies like **AWS, GCP, Azure** and **IBM Aspera Cloud**.
* Expertise in creating **executive Tableau Dashboards** for Data visualization and deploying it to the servers; also skilled with **Power BI, AWS Quick Sight, seaborn and matplotlib.**
* Working knowledge of Database creation, **CRM** operations, **ETL** tasks using **Informatica Power center**, Integration and writing complex SQL queries in **Microsoft SQL server**.
* Experience in converting the businessspecifications into technical requirements, creating various Static, Ad-Hoc, Technical Reports, **business KPI’s,** self-serve visualizations and articulating it across all the stakeholders.
* Experienced in running stakeholder meetings, interacting with **business stakeholders**, using Project Management tools like **MS Teams, SharePoint,** **JIRA,** and working in an **Agile** Methodology.

**SKILLS**

|  |  |
| --- | --- |
| **Languages** | Python, SQL, R, MATLAB |
| **Database** | AWS S3, AWS Athena, AWS Redshift, MySQL, Microsoft SQL Server, Spark SQL. |
| **Statistical Tests** | ANOVA tests, t-tests, z-test, Chi-Square Fit test, Multivariate tests. |
| **Validation Techniques** | Monte Carlo simulations, k-fold cross validation, A/B Testing. |
| **Optimization Techniques** | Grid Search, Proximal Gradient Descent, Stochastic Gradient Descent. |
| **Statistical Methods** | **Regression Methods:** Linear, Polynomial, Lasso & Ridge Regression.  **Classification Methods:** Logistic Regression, K-NN, Decision Trees, Naïve Bayes, Support Vector Machines (SVM).  **Clustering:** Exclusive Clustering (K-means), Hierarchical Clustering, Overlapping Clustering (C-means), Association Rule Mining.  **Ensemble Learning:** Random Forest Models (Bagging & Boosting).  **Dimensionality Reduction:** Feature Selection (exhaustive search, SFS, SBS, SFFS, SBFS, +l -r) and Feature Transformation Techniques (PCA, LDA & MDA).  **Time Series:** Stochastic Signal Processing, Estimation Theory, AR, MA, ARMA, ARIMA, ARCH & GARCH, RNN based Deep Learning models like DeepAR. |
| **Data Visualization** | Tableau, Power BI, AWS Quick Sight, matplotlib, seaborn, MS Excel, MS Visio. |
| **Big Data** | Apache Spark, Hadoop, PySpark, AWS – Sagemaker, Lambda, EMR, Step Functions |
| **Cloud Services** | AWS, Azure, GCP, IBM Aspera Cloud. |
| **Related Courses** | Pattern Recognition & Analysis, Estimation Theory, Stochastic Signal Processing. |

**EDUCATION:**

**Michigan State University East Lansing, MI Master of Science, Electrical & Computer Engineering August 2017 – May 2019**

**Amrita University Kollam, India Bachelor of Technology, Electrical Engineering 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>

**EXPERIENCE**

**Microsoft – HCL Chicago, IL Data Scientist March 2021 - present**

**Description:** As part of Field Marketing & Operations Team, was involved in Forecasting & Optimization projects such as **Resource Utilization Forecasting** and **Strategic Capacity Planning Optimization**.

**Responsibilities:**

* Extracted data from On-premises database using SQL queries in **Microsoft SQL Server** Environment.
* Analyzed the data and handled anomalies in the dataset by statistical methods.
* Forecasted resource utilization using ML algorithms like **Auto ARIMA**, Holtwinters and **FB Prophet**.
* Compared the forecasts and chose the best forecast for each cross-section in the dataset.
* Analyzed the resource capacity planning use case and performed literature review to suggest a capacity planning model.
* Handled anomalies in the capacity planning dataset by an algorithmic approach and business logics.
* Performed Data Analysis on and tested different capacity models formulated by the team.

**Environments:** Python, SQL, Microsoft SQL Server, Excel, Power BI.

**Koch Industries Chicago, IL Data Scientist / Data Science Analyst July 2019 – February 2021**

**Description:** This project required me to improve the **Demand Forecasting** accuracy of the SKU’s belonging to different Business Units and manufactured by Molex, Koch Industries company by following **CRISP-DM** methodology and using Advanced Time Series ML models; provided business strategy recommendations; have setup reporting solutions across multiple business areas and reported supply chain KPI metrics and value add created using the ML models to the stakeholders. The adjusted SKU’s Demand Forecasts are then used down the **Supply Chain** to adjust the inventory levels in the warehouses.

**Responsibilities:**

* Gathered **business requirements,** converted them to technical requirements and worked with **cross functional stakeholders** to understand existing supply chain process flows and to provide business recommendations, display root cause analysis of the problems, identify the areas of improvement in current operations, business KPI’s and to point key zones to be tracked for driving business outcomes.
* Performed **Business process modelling** (BPMN) in **MS Visio** inorder to analyze and automate the existing practices. Have run the stakeholder meetings in the absence of Project Manager.
* Extracted data at SKU and Customer level from AWS **Athena** using SQL queries; Explored and manipulated large **RDD Datasets** from legacy Hadoop cluster using **PySpark**.
* Munged the data to several formats inorder 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** &checked for **descriptive** & **prescriptive** statistics.
* Have done **macroeconomic analysis** for the SKU’s based on End Customer, Location - Region & Country and have come up with around 130 external features like Consumer Price Index, Producer Price Index, Annual Sales, Interest Rates, GDP, Consumer Confidence Index…etc.
* 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.
* 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 suitable prediction model.
* Applied 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 inorder to drive key business decisions.
* Calculated error metrics like **MAD,** **WMAPE** at lag, customer, plant, region levels for the models using the generated forecasts and conducted **root cause analysis**, compared the KPI’s and error metrics with baseline models, reported accuracy differences and communicated the best models to implement in production to the stakeholders.
* Used **Tableau** and **AWS QuickSight** to present insights and business strategy recommendations to executives and stakeholders in the form of user stories and dashboards.
* Assisted the deployment team to integrate the finalized forecast models with **SAP APO** and configured 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 QuickSight, AWS Sagemaker, AWS S3, AWS Athena, AWS Lambda, Glue, Gitlab, Excel, Visio.

**General Motors Lansing, MI Data Science Analyst August 2017 - May 2019**

**Description:** This project required me to analyze the **CAN (Controller Area Network)** sensors data and predict the future status of the car using stochastic and time series ML algorithms; also required me to analyze **Power Train Control Module (PCM)** data and find the relationship between **Diagnostic Trouble Codes (DTC)** and the Engine Sensor Data of the car and to develop an **API** that performs Real Time Data Analytics on the input data, and predicts the reason and exact sensors responsible for DTC’s.

**Responsibilities:**

* Gathered business requirements, converted them to technical requirements and communicated with stakeholders for better understanding of the process flows and areas of improvement.
* Logged and Stored the 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).**
* Performed **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 inorder 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 inorder 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** 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 analysis in **PyCharm** using python inorder 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 **Power BI** and are presented to the stakeholders.

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

**TATA Consultancy Services Pune, India Assistant Systems Engineer / Data Analyst July 2016 – July 2017**

**Description:** My engagement with TCS has given me a chance to configure **Manufacturing Execution System (MES)** for the production plants of P&G across the world; to analyze the manufacturing plant data, produces **ad hoc reports** and **business recommendations** to senior managers inorder to reduce downtime and thus increasing the **value add** to the organization.

**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 Kerala, India   
Intern / Data Analyst December 2015 - May 2016**

**Description:** This project required me to design and develop a PLC based **Automated Grain Dispensing System** inorder to automate the ration system in rural areas of India; to **analyze** the PLC data for sensor calibration and recommend ways to improve the **efficiency** of the system.

**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 trails 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.