**ATUL KUMAR SHARMA**

[sharmaaks50@gmail.com](mailto:sharmaaks50@gmail.com) | <https://www.linkedin.com/in/>atul-kumar-sharma-78181941 | [Github](https://github.com/atulsharma89/)|+1(408) 816-4288

**EDUCATION**

**The University of Texas at Austin, McCombs School May 2021**

PG Program in Artificial Intelligence and Machine learning **GPA: 99.38/100.0**

**SRM University, Chennai May 2012**

Bachelor of Technology: Electronics & Instrumentation Engineering **GPA: 8.9/10.0**

**SKILLS**

**Programming :** Python, SQL, Java

**Data Engineering :** Teradata, PostgreSQL, SQL Server, MySQL, Oracle, MongoDB, Snowflake

**Tools :** GitHub, Docker, JIRA, Visual Studio, Power BI, Tableau, Databricks, Eclipse, Jenkins, Alteryx

**Big data & Cloud :** Hadoop, Spark, Hive, AWS (S3, EC2, IAM, Data Lake, Auto ML), Google Cloud (Kubernetes)

**RELEVANT EXPERIENCE**

***Data Scientist,* *AppleCare Business Insights (Apple), Sunnyvale, CA*** *#Data Science #ML* **May’19 – Current**

* Built a regression-based Machine learning model using Statistical tools and ML algorithms like XG Boost and Random Forest with 86% accuracy to predict customer turnaround time for global Mail-In repairs to measure Repair Centre performance.
* Built a Python based **Text Extraction System**, to extract the **correspondence message** from Sonar toolfor tracking retail escalations (Returns, inventory, back ordered, DOA parts) that helped in increasing AppleCare Buyers efficiency by 50%.
* Built a SQL/Tableau based **Repair Turn Around time tracking data solution** that helps 400+ global users to track the state of open/closed repairs to identify the key potential drivers to improve the customer’s repairs experience by more than 60%.
* Built an **Alteryx** based **automated data induction and reconciliation framework** to induct & validate data from various sources and setting up complex DQM to track and maintain health of growing data.
* Built an **ANN** & **Logistic regression-based** classification model using **Python** **Keras** for identifying AppleCare contract churn rate for **iPhone** contracts to reduce the churning rate of nearly .5 million customers yearly with 85% accuracy.
* Built Python and Kubernetes based multi layered application for helping Demand Planning and Supply Chain teams to get a 360-degree view of BOM data for parts Forecasting and last-time-buy planning for apple parts.

***Data Science Engineer,* *TCS, client: Apple,*** *#ETL #Data Engineering #Data Analysis* **Sept’12 – May’19**

* Ingested data from various sources through SQL, AWS data lake, Box, snowflake using Python to create data views that are consumed by users for **parts pricing analysis** and catching **price anomalies** which resulted in saving $1M annually.
* Built **Python/Teradata/Tableau** based end-to-end data centric solution to predict and analyze health of **4 hours repair SLA** for enterprise customers which helps to serve more than 20000 AppleCare Enterprise customers across the globe.
* Designed and built **statistical analysis** models on large factory datasets that helped users to catch parts health anomalies.
* Formulated next generation analytics environment, providing self-service, centralized platform for all data-centric activities which allows full 360-degree view of **customers repair experience** with Apple through various channels.
* Collaborated cross-functionally with business analysts, developers, and testers to explain new process transformations.

**OTHER RELEVANT PROJECTS**

***Credit Card Fraud Detection,*** *#Anomaly detection #Feature Engineering #Neural network*

* Used SMOTE to overcome class imbalance and PCA & SVD to reduce feature dimensions. Supervised: Random Forest, Neural networks with 98% accuracy. Unsupervised: segmentation & K-means clustering used to recognize fraud patterns.

***AppleCare Database modelling, #****Data Engineering #ERD #ETL #SQL #StoredProc #Partitions*

* Developed a Database for a AppleCare parts mapping use case, with ERD modeling to meet specific business needs.

***Customer Churn Modelling,*** *#Marketing Analytics #Feature Engineering* ***#****Predictive modelling*

* Explored and preprocessed the Telco Churn Dataset using pandas to predict churn. Tuned hyperparameters, did randomized search & feature importance to improve performance. Evaluation metrics: AUC/ROC, Recall, Precision and F1 score.

***Netflix - A Movie Recommendation System, #****Machine Learning* ***#****Recommendation engine*

* Implemented a hybrid recommendation system using Collaborative filtering and Pearson’s R correlation for suggesting movies for users based on prior user ratings. Achieved recommendations identifying pairs using Jaccard similarity.

***Sentiment Analysis & Neural Machine Translation, #****NLP #Artificial Intelligence #Social media data*

* Implemented a text classifier which sorts & tags emotions using Naïve Bayes algorithms by likelihood and frequency.

***Lending Club – Loan Default Prediction,*** *#Machine Learning* ***#****Pandas*

* Extracted, Pre-processed, and manipulated data from Lending club. Built Predictive models such as Gradient boost and SVM.