Rupam Jogal

Sunnyvale, CA | ☎ (857) 333-7822 | ✉ rupamjogal@gmail.com
www.linkedin.com/in/rupam-jogal-0712 | Online portfolio: https://rupam07.github.io/

SKILLS

Programming languages	R, Python (numpy, pandas, seaborn, sklearn, matplotlib, pickle), HTML, CSS, SQL, C++, Exasol
Databases/ETL/Cloud Platform	SQL Server, MySQL, Spark, Cloudera, AWS (S3), SSIS, Alteryx, Hive, Databricks, DBeaver
Frameworks/ Platform's	PyCharm, Tableau, Microsoft PowerBI, Advanced Excel (vlookup), Jira, Confluence, Sublime
IDE'S	R-Studio, Matlab, MS Office, Latex, Visual Studio, Scilab, NetBeans IDE, Eclipse, GitHub, Jupyter, Postman

EDUCATION

Northeastern University | Boston, USA | MS in Analytics – GPA: 3.75 **Apr 2020**

- **Relevant Coursework:** Probability Theory & Statistics, Data Mining Applications, Data Visualization, Predictive Analytics, Data Management and Big Data, Application of AI, Data Warehousing and SQL, Data-Driven Decision Making

Ahmedabad University | Ahmedabad, India | Bachelor of Technology in Information and Communication Technology **May 2018**

- **Relevant Coursework:** Object-Oriented Programming, Data Structures and Algorithms, Machine Learning

EXPERIENCE

Aera Technology | Mountain View, CA | Associate Skills Developer **Dec 2020 – Present**

- Building **Cognitive Data Layer** (CDL) that works for different customer **ERP** systems, as a base for building skills on top that would help in predictive analytics and report creation
- Design and build cognitive application using **Exasol** specifically for SAP and JDE Oracle ERP by **loading the data** first to data crawler that **extract** new as well as delta changes, **transform the data** using Aera's ETL platform and troubleshoot issues during data load and processing
- Creating a **harmonized datasets** to store and connect both ERPs to Aera Model by implementing complex SQL queries to store the data
- Implementing **data warehousing techniques** to create dimension and facts links, validating the implemented models and optimizing the code for good data quality, creating measures for numeric datasets that implement the major quantity and amount calculations
- Performing data analysis on the base layer to check the accuracy of the implementation for complex use cases and **generating reports** specific to the customer data and build **KPIs**/attributes required to visualize customer data

Aera Technology | Mountain View, CA | Data Science Intern **Jun 2020 – Dec 2020**

- Building cognitive skills to help Pharma, CPG customers in predicting unfulfilled demand to market affiliates using lead-time values based on history data, extracting features and grains values with **Microsoft IDEAR** library increasing prediction accuracy by 90%
- Performing clustering analysis for Pharma client, discovered 6 groups through unsupervised machine learning model, predicting lead-time values using gradient boosting, deploying it using backend **API**s in **Postman** and Python with model accuracy of 97%
- Building forecasting model using predictive analytics on GPU with **xgboost** and random forest regressor which resulted in 3.8% AUC lift, aggregating and preprocessing large amount of data for CPG client with factors like location, weather, SKUs, volume, sales etc
- Applying feature importance techniques using **shap values** and xgboost **feature importance** to get the important features affecting the lead-time value to present as POC for client
- Implementing **Agile Scrum Software Development Lifecycle** methodology in preparing test cases and monitoring the performance of the Epic and user stories using **Jira** to report and test the production environment

Northeastern University | Boston, MA | Customer Service Analyst **Dec 2018 – Mar 2020**

- Managed more than 200 students, 15 high profile faculty for providing them unrestricted access round the clock to the knowledge treasure
- Performed **ETL** using **Alteryx** and generated **dashboards** using Tableau which guided in checking the available day to day inventory using the dataset gathered in the past with 87% accuracy delivering accurate time frame to students and faculty to pick up electronics
- Designed a Tableau frontend for incoming calls and wait calls, conducted data **quality assurance** used for improving the hardware and software technical support provided in person
- Troubleshoot an issue and raised over 1000 tickets using **ServiceNow** and aided them resolve over the call

Pixometry Infosoft Pvt. Ltd. | Ahmedabad, India | Business Data Analyst **Jan 2018 - May 2018**

- Applied topic modeling using **Gensim** to check the product quality based on suggestions. Built a dictionary and a corpus of over 7000 words using Python which helped in counting the words in each file, creating a **Term-Frequency – Inverse Document Frequency**
- Tokenized words further created the bigrams and trigrams from the words and performed the topic modeling with unsupervised machine learning **Latent Semantic Indexing** (LSI) algorithm which used **SVD**, summarizing the percentage contribution of each word in the files using a pre-built **Word2Vec** model and hence gave a prediction accuracy of **68%**

PROJECTS

Burial Record Image/Text Recognition **Dec 2019**

- Applied classification using **keras** and **tensorflow** in backed to the dataset with dead burial records using max-pooling and relu and classified them into 6 classes with **convolution neural network (CNN)** achieved an accuracy of **94%**
- Recognized texts with Optical Character Recognition using **opencv pytesseract** and then **AWS textract** and achieved a confidence interval with AWS and **S3** for all the images in the dataset around **95%**

Customer Churn Analysis **Aug 2019**

- Deployed a model in order to predict the churn percentage in the dataset based on 10 factors such as customer usage patterns and analysed using a classification machine learning algorithm across churn rate
- Classified using **sci-kit** learn **RandomForestClassifier**, **Naïve Bayes** (GaussianNB), **kNN** (KNeighborsClassifier) and **LogisticRegression**, evaluated confusion matrix scored for every model with random forest model with the 97% accuracy among all other models

Database Design - Online Shopping Management **Oct 2018**

- Created normalized **ER model** using **Visio**, implemented check constraints, stored procedures, views column encryption and SQL queries that answering question about database in **MySQL**, generating reports based on the quality, price and geography on **PowerBI** with visualizations