Rupam Jogal

Boston, MA | ☎ (857) 333-7822 | ✉ jogal.r@husky.neu.edu
www.linkedin.com/in/rupam-jogal-0712 | github.com/Rupam07

SKILLS

Programming languages	R, Python (numpy, pandas, seaborn, sklearn, matplotlib), HTML, CSS, C, C++, SQL
Databases/ETL	SQL Server, MySQL, Spark, Cloudera
Machine learning algorithms	Regression Analysis, Classification, Decision Tree, Forecasting, Support Vector Machines, KNN
Business intelligence tools	Tableau, R-shiny, Microsoft PowerBI, Advanced Excel, Qlik View, Qlik Sense
Operating systems	Windows, Linux, Mac
IDE'S	R-Studio, Matlab, MS(Word, PowerPoint, SharePoint, OneDrive, Excel (vlookup)), Latex, Visual Studio, Scilab, NetBeans IDE, Eclipse, GitHub, Jupyter, Hadoop, AWS
Communication skills	Excellent written and oral communication skills

EDUCATION

Northeastern University | Boston, USA **Apr 2020**

*MS in Analytics – **GPA: 3.88***

- **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 **May 2018**

Bachelor of Technology in Information and Communication Technology

- **Relevant Coursework:** Object-Oriented Programming, Data Structures, and Algorithms, Linear Algebra, Database Management Systems, Operating Systems, Machine Learning

PROJECTS

Image recognition **Oct 2019**

- Predicting logos of famous brands using keras and tensorflow in Python while implementing part localization to resize the images, encoding logos in binary format and the labels with one-hot encoding for preprocessing
- Implementing convolution neural network with max pooling and reducing the parameters progressively to control overfitting, using tanH and ReLU activation in CNN increased the accuracy by 10%

Customer Churn Analysis **Aug 2019**

- Deploying a machine learning-based classification model in order to predict the churn percentage in the dataset based on the customer usage patterns and churn data whether the customer has churned or not
- Performing analysis using a classification machine learning algorithm for figuring the 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 best accuracy among all other models

Sentiment Analysis **Jun 2019**

- Implementing Natural Language Processing on the dataset with textual requests for Pizza, tokenized and lowercased the sentences and words in the data and applying the statistical model to gain information from data.
- Identifying the stop words as well as punctuations and removing them to only have a gist of the document. Stemming and Lemmatizing words with the help of Parts of Speech tags and applying tangential note on the respective lemmatization. Vectorizing the words with sci-kit learn
- Applying Naïve Bayes classification and SVM to find out if the requester gets the pizza or not and see is their review is positive or negative

Pest and Disease identification in plants **May 2018**

- Identifying different stages of disease in the cotton plant, converting the RGB image into grayscale to do the leaf edge detection by removing the background noise and reducing the image pixels
- Segmenting the leaf image with k-means clustering and extracting features like texture for the classification of diseases with the neural network classifier with an accuracy of 77.8%

Facial Expression Recognition **Apr 2017**

- Identifying different facial expression of a given image like angry, sad and happy using viola jones face detection algorithm, extracting the haar features using the convolution kernel which helps in detecting the presence of a feature in the image
- Applying machine learning algorithm AdaBoost to help to find the best feature with an accuracy of 82%. Finally, a cascading classifier is used to group all the features in several stages and each has a certain number of features in it

EXPERIENCE

Pixometry Infosoft Pvt. Ltd. | Ahmedabad, India **Jan 2017 - May 2017**

Business Data Analyst

- Applying topic modeling with the help of Gensim in order to help in the risk prediction. Built a dictionary and a corpus using Python which helped in counting the words in each file, creating a Term-Frequency – Inverse Document Frequency
- Tokenizing 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 helped in providing the percentage contribution of each word in the files using a pre-built Word2Vec model and hence gave a prediction accuracy of 68%