ROHIT RAJESH KHATU.

Boston, MA | (617) 818-3265 | github.com/rkhatu97 | khatu.r@northeastern.edu | linkedin.com/in/rkhatu97

EDUCATION

Northeastern University, Boston, MA **January 2020 – July 2021**
Master's in Data Analytics with concentration in Applied Machine Intelligence.
Current GPA: 3.93/4.0.
Related Courses: Probability Theory and Statistics, Big Data, Data Visualization, Intermediate Analytics, Data Mining.
Saraswati College of Engineering, Mumbai University, Maharashtra **June 2014 - May 2018**
Bachelor of Engineering, Computer Engineering
Related Courses: Operating Systems, Big Data, Soft Computing and Data Structures.

WORK EXPERIENCE

Machine Learning Engineer Intern, Dtonomy Inc. (Client: Columbia University) **January 2021 – June 2021**

- Refined random forest model used for pattern discovery by performing feature addition and hyperparameter tuning. Deployed the model as a service on Docker container for identifying security detection and provide analysis.
- Constructed a node-red workflow using NodeJS, JavaScript, html, and python to set inbox rules to block spam mails with key word and to block sender with id, leveraging Microsoft Graph API and Azure app registration feature. Wrote selenium script to automate the process for requesting and receiving consent for scopes of the graph API used to access and set the inbox rules.
- Created a PowerShell node-red workflow to run PowerShell script and perform activity to set enterprise level mail flow rules to block spam mails and to block sender over Microsoft Exchange Server. Achieved OS independency for the node by installing PowerShell dependencies into docker container and deploying the node on top of container.

Associate Consultant (DevOps Developer), Mindcraft Software PVT.LTD. **August 2018 - December 2019**

- Leveraged DevOps technology; Jenkins, docker, and chef for automated deployment of MySQL server and client on multiple servers simultaneously which reduced the deployment time to 2 Hours from 1 week. Deployment of Loan disbursement Web application on Tomcat server was automated for DEV/UAT/PROD environment using single Jenkins job.
- Wrote Shell scripts to perform database/WAR file backup, user/database/table creation and set user access rights as per project security guideline. Triggered these shell scripts using automation tools.
- Engineered software with help of ruby scripting for Universal Overseas Bank, documented POC's for entire project, worked under Agile practices. Working Knowledge of finance and insurance sector.
- Deployed on overseas project at client site in Singapore, achieved timely deployment and testing**.**

TECHNICAL SKILLS

Programming Languages: SQL, Java, Python(NumPy, Pandas, SciPy), R(dplyr, ggplot2), JavaScript, Shell Scripting.
Web Development and Web Server: HTML, CSS, Bootstrap. Apache Tomcat.
Database: MySQL, MariaDB.
Cloud Technology and Automation: Docker, AWS (EC2, S3), Azure, GCP, Jenkins, bamboo, Jira, chef.
Tools and Packages: GIT, Linux, Microsoft Office, Tableau, PowerBI, Anaconda, TensorFlow, PyTorch, NodeJS.
Big Data Technologies : Hadoop, HDFS, Sqoop, Hive, MapReduce, Spark.
Statistics : Hypothesis Testing, Descriptive Statistics, A/B Testing, ANOVA, Chi-Square Test, Pearson Test.
Machine Learning Algorithms : Linear Regression, Ridge and Lasso Regression, Logistic Regression, KNN, Naïve Bayes, K-Means clustering, Apriori, Decision Tree, Random Forest, Boosting Algorithms, PCA, SVM, CNN, LSTM.

PROJECTS

Housing Sales Analysis: Tableau , Python **November 2020**

- Imported the data into python to perform EDA, removed anomalies and irregular data. Performed data mutation to fill out the null data based on interdependency of columns. Achieved Data cleaning by inspecting and removing remaining null values and by eliminating unwanted symbols throughout the dataset. Exported the dataset in csv.
- Imported the cleaned dataset to leverage the power of tableau to perform data analysis and data visualization by building and deploying dashboard using line charts, pie charts and bar plots to show difference in sales prices based on year built, heating amenity and roof and house structure. Helps in investing in house which gives maximum ROI.

Product Review Sentiment analysis: Python, HTML, CSS (Flask Deployment) **September 2020**

- Analyzed 20 years of amazon product review, by segregating reviews based on ratings that is below 2 stars as negative, equal to 3 stars as neutral and above 3 as positive, built bar plot to visualize data count of all three sentiment.
- Constructed ML model by first transforming the data using TFID vectorizer and applying Naïve Bayes, logistic regression, and Linear SVM in python, choose logistic regression as best of all based on metrics, saved model as PKL file and used to accurately predict sentiment of amazon reviews to see if a product has successfully captured the market.

Crimes in Boston Data Analysis: R, RShiny **February 2020**

- Analyzed a million entries dataset on crimes happened in Boston from year 2016-2019. Develop heat map to show increase and decrease in crime rate as per county per year.
- Used ARIMA forecasting model to predict the change in crime rate for six months.