EASHAN ARORA

Chicago, Illinois| eashanarora11@gmail.com | (312) 391-1892
Linkedin | Project Portfolio | Github

EDUCATION

University of Illinois Chicago **Aug 2021-Dec 2022**
Master of Science in Business Analytics **GPA:3.78/4.0**

- Relevant Coursework: Statistics, Data Mining, Machine Learning, Analytics for Big Data

Jaypee Institute of Information Technology, Noida, India **Jul 2016-Jun 2020**
Bachelor of Technology in Computer Science

- Relevant Coursework: Software Engineering, Computing for Data Science, Database Systems

WORK EXPERIENCE

Lumiq
Data Scientist **Jul 2020-Jul 2021**

- Created Interaction Engine for an insurance Firm in India using an open source chatbot platform Rasa resolving 5000 customers queries coming in form of human utterances
- Customized a RASA-based end to end machine learning pipeline, assessed use of 6 modules proposed in conversational AI and annotated users responses to improve training by 10%
- Designed advanced SQL scripts to identify anomalies and perform data validation on 1M+ records resulting in better and faster query performance
- Collaborated with team for a Video/Image Analytics project to extract relevant information from 100+ legal documents employing machine learning frameworks reducing human effort by 40%

Radix Info Solutions **Jun 2018-Aug 2018**
Data Analyst Intern

- Executed data mining techniques along with data cleaning, visualization and executed statistical test methods for calculating significance of each attribute compared to target variable in a particular dataset of 10000 instances
- Prepared tableau dashboards to synthesize data into reporting formats and deliver recommendations to client

TECHNICAL SKILLS

- **Programming Languages** : Python, R, SQL, Spark, MySQL, Postgres, SQL Server, Oracle
- **Machine Learning and Statistics Techniques**: Regression, Clustering, Classification, Data modeling , EDA, Bagging (Random Forest), NLP, K-NN**,** Hypothesis Testing, T-test, Anova, Chi-Squared
- **Tools**: Tableau, PowerBI, AWS, Microsoft Excel, Jupyter Notebook, Rstudio, Github
- **Libraries:** Pandas, Numpy, Scikit learn, ggplot, Matplotlib, dplyr, Seaborn, Plotly

ACADEMIC PROJECTS

Natural Language Understanding Pipeline (Python, Transformers, NLP)

- Developed an automated NLU pipeline including 5 modules aiming to provide a semantic way for user utterances
- Performed intent classification (context of sentence), entities extraction (keywords of user utterances), query type, correctness (probability for input being correct) and sentiment analysis for any user utterance through NLP
- Obtained Probability scores for models utilizing simple transformers framework greater than 95%

Jetblue Airlines Performance Dashboard (Tableau, Microsoft Excel)

- Constructed a Tableau dashboard for Jetblue Airlines 2020 dataset to identify flight delay patterns and suggest solutions to improve efficiency. For example, analyzing reasons that caused 51% of delays
- Evaluated strengths and shortcomings by comparing derived metrics with 10 other airlines and suggesting additional refinements.

Predicting Employee Attrition in Recession (Python, Classification)

- Implemented 3 classification models comprising decision tree, random forest, logistic regression to predict attrition and facilitate decision making for human resources department
- Prepared Exploratory Data Analysis, advanced data cleaning ,feature engineering, feature encoding, attribute selection to examine factors behind attrition
- Achieved over 80% accuracy scores for all models on an extremely sparse dataset of 20000 rows

RESEARCH PUBLICATIONS

- E.Arora, S.Mishra, K. Vimal Kumar, P.Upadhyay, “Extending Bidirectional Language Model for Enhancing the Performance of Sentiment Analysis”, published at International Conference on Cybernetics, Cognition, Machine Learning Applications 2019 (Springer)