Kratika Agrawal

(774) 701-9288 | United States
kagrawal@wpi.edu | linkedin.com/in/kratika-agrawal-wpi | kratika1008.github.io/portfolio/

EDUCATION

Worcester Polytechnic Institute (WPI) Worcester, MA, USA
Master of Science in Data Science, GPA – 4.0/4.0 *Aug. 2019 – May 2021*

- **Related Courses:** Statistical Methods for Data Science, Machine Learning, Intro to Data Science, Big Data Management, Deep Learning, Information Retrieval & Social Web, Data Mining Business Applications, Predicting Human Decisions

Madhav Institute of Technology & Science Gwalior, MP, India
Bachelor of Engineering in Electronics, GPA – 8.62/10 *Aug. 2010 – May 2014*

TECHNICAL SKILLS

Languages/Framework: Python, Java, Scala, SQL, JavaScript, R, Linux, Hadoop, Apache Spark, MongoDB
Python Libraries: Tensorflow, Keras, Pandas, NumPy, Matplotlib, SciPy, Scikit-learn, Plotly, NLTK
Tools: Git, Google Cloud Platform, IBM SPSS Modeler, Anaconda, VS Code, IntelliJ, Eclipse, Excel, ANT, JIRA, Salesforce

WORK EXPERIENCE

Graduate Research Assistant | *Worcester Polytechnic Institute, USA* Sep 2020 – Present

- Developing new state-of-the-art model to generate conditional Text Sequences using GAN and evaluate its performance
- The goal is to use the sequence generator to augment conditional texts to train model for making diagnostic predictions
- Continuously doing literature review and evaluating related work in the field of sequence generation

Data Science Intern - GQP (Capstone) | *Basis Technology, USA* Aug 2020 – Dec 2020

- Worked on improving the Entity Disambiguation task of the Natural Language Processing for the Entity Linking process
- Designed a novel model using graphs and CNN to perform disambiguation more accurately using AIDA-CoNLL dataset
- Actively participated in brain-storming sessions with company's mentors and product manager

AI/ML & Big Data Research Intern | *Futurewei Technologies, Inc., USA* May 2020 – Dec. 2020

- Contributed to the open-source Apache Spark project to optimize performance up to 20% of the current shuffle stage
- Implemented aggregate-based shuffle in Spark using storage systems like Alluxio and HDFS and evaluated it
- Modified the code for reads in merge-based shuffle implementation using cache memory
- Automated the process to run experiments and collect metrics in CSV format by creating custom shell scripts

Salesforce Associate Developer | *Cognizant Technology Solutions, India* Feb. 2018 – May 2019

- Optimized and enhanced the search in a directory portal for UK financial authority in as fast paced agile project
- Visualized various results by creating custom Reports and Dashboards with complex functions
- Written test classes to make code production ready with up to 90% code coverage

Senior System Engineer | *Infosys Limited, India* Jul. 2014 – Jan. 2018

- As Siebel CRM developer, saved >90% time by automating Exchange & Return process integrating many end-system APIs
- Closely worked with stakeholders and business analysts on client-site to understand and deliver requirements
- Provided project training sessions to new joiners and mentored them throughout the project
- Supported during production release and post-production to provide quick bug-fixes

ACADEMIC PROJECTS

Visual Question Answering (VQA), WPI | *Object Detection, LSTM, CNN* Mar. 2020 – May 2020

- Achieved 64% accuracy on Image COCO dataset in selecting the topmost answer amongst the voted answers lists
- Retrieved image features using pre-trained ResNet-50 and Region Proposal Network output of faster RCNN
- Used GloVe embedding for training textual questions and answers

Customer Churn Prediction, WPI | *Python, Scikit-learn, Matplotlib, Seaborn* Mar. 2020 – Apr.2020

- Gained AUC score of 0.826 in predicting customer churn on a skewed Kaggle dataset using Ensemble classification model
- Analyzed the data and performed data mining to better understand the customer and various features affecting the churn
- Developed metrics and visualized the data and hypothesis result to recommend ways for customer retainment

Predicting Patient's Survival, Kaggle Competition | *Scikit-learn, Seaborn* Jan. 2020 – Feb. 2020

- Achieved 0.909 AUC-ROC curve score in predicting patient survival using first 24 hours ICU data with Light-GBM
- Extracted important features among 185 total numerical and categorical features using Random Forest
- Trained the model with Stratified-KFold and tuned hyper-parameters using grid-search for the probability regression

Big Data Hadoop & Spark Implementations, WPI | *Java, Scala, Apache Spark* Aug. 2019 – Dec. 2019

- Saved up to 60% time in performing join on large, skewed, distributed dataset in Spark compared to simple join
- Performed spatial join of data points in x-y plane and rectangles in the same plane using Hadoop & Spark
- Implemented k-means clustering on custom generated 2D coordinate points on distributed network using Hadoop

AWARDS AND ACCOMPLISHMENTS

- Received Best Teamwork Award for the capstone project (GQP) Dec 2020
- Won Best Project award for Deep Learning class final project (VQA) May 2020
- Ranked top 10% in WiDS Datathon 2020, Kaggle (2nd Runner-up among WPI teams) Feb. 2020
- Insta-Award winner at Infosys Limited for exemplary performance May 2016