

# Kratika Agrawal

(774) 701-9288 | 231 Russell St, Worcester, MA, USA 01602

[kagrawal@wpi.edu](mailto:kagrawal@wpi.edu) | [linkedin.com/in/kratika-agrawal-wpi](https://www.linkedin.com/in/kratika-agrawal-wpi) | [kratika1008.github.io/portfolio/](https://kratika1008.github.io/portfolio/)

## EDUCATION

### Worcester Polytechnic Institute (WPI)

Master of Science in Data Science, GPA – 4.0/4.0

Worcester, MA, USA

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

Bachelor of Engineering in Electronics, GPA – 8.62/10

Gwalior, MP, India

Aug. 2010 – May 2014

## TECHNICAL SKILLS

**Languages/Framework:** Python, Java, Scala, SQL, JavaScript, HTML, R, Apache Pig, Hadoop, Apache Spark, MongoDB

**Python Libraries:** Tensorflow, Keras, Pandas, NumPy, Matplotlib, Scikit-learn, NLTK, SciPy

**Tools:** Git, Google Cloud Platform, Anaconda, VS Code, IntelliJ, Eclipse, Excel, ANT, JIRA, Salesforce

## ACADEMIC PROJECTS

### Capstone Project for Entity Linking (GQP), Basis Technology | Python, Tensorflow

Sep. 2020 – Dec. 2020

- Worked on improving the Entity Disambiguation task of the Natural Language Processing for the Entity Linking process
- Developed 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

### Visual Question Answering (VQA), WPI | Object Detection, Keras, 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 | R, Python, Scikit-learn, Matplotlib

Mar. 2020 – Apr. 2020

- Gained AUC score of 0.826 in predicting customer churn on a skewed Kaggle dataset using Ensemble classification model
- Performed data cleansing, data exploration, feature extraction and sampling on the dataset
- Visualized the data and relationship of churn rate to key features, also summarized results and ways to reduce the churn

### Predicting Patient's Survival, Kaggle Competition | Scikit-learn, Seaborn

Jan. 2020 – Feb. 2020

- Achieved 90.914% AUC-ROC curve score in predicting patient survival using first 24 hours ICU data
- Extracted important features among 185 total features using Random Forest
- Trained the dataset using Light-GBM with Stratified-KFold of 5 folds and tuned hyper-parameters using grid-search

### 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

## DIRECTED RESEARCH

Sep 2020 – Present

- Developing new state-of-the-art model to generate conditional Text Sequences and evaluating its performance
- The goal is to use the sequence generator to augment conditional texts to train model for diagnostic prediction

## WORK EXPERIENCE

### AI/ML & Big Data Research Intern | Futurewei Technologies, Inc., USA

May 2020 – Dec. 2020

- Worked on the open-source Apache Spark project to optimize performance up to 20% of the current shuffle stage
- Created shell scripts to automatically run experiments and collect metrics in CSV format
- Worked on merge-based shuffle implementation to modify reads using cache memory
- Implemented aggregate-based shuffle in Spark using storage systems like Alluxio and HDFS

### Salesforce Associate Developer | Cognizant Technology Solutions, India

Feb. 2018 – May 2019

- Created module to predict sales graph and probability for lead conversion using Salesforce Einstein.
- Worked on agile project to optimize and enhance the search directory portal for UK financial authority
- Led the project to migrate code base from one salesforce organization to another
- 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
- Worked on client-site closely with stakeholders and business analysts 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

## 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