

HARSHIT GUPTA

Champaign, Illinois - 61820 | hrgupta@outlook.com | +1 217-418-7056 | [LinkedIn](#) | [GitHub](#)

EDUCATION

School of Information Sciences, University of Illinois at Urbana-Champaign, USA Candidate for Master of Science in Information Management.	May 2020
Thadomal Shahani Engineering College, University of Mumbai, India Graduated as Bachelor of Engineering in Computer Engineering with First Class.	May 2016

SOFTWARE SKILLS

- Programming Languages/Tools: Java, Python, MySQL, R, HTML, CSS, JavaScript, Boto3, Git, Docker
- Cloud Technologies: AWS EC2, S3, VPC, ECS, RDS, Spark MLlib
- Libraries: NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow, Keras, Spacy
- Data Science Platforms: Jupyter Notebook, RStudio, Tableau, AWS SageMaker, Google AutoML

PROFESSIONAL EXPERIENCE

Synchrony Financial, Champaign, USA <i>Data Scientist Intern – Emerging Technology</i> <ul style="list-style-type: none">• Developed an ML solution to find suspicious merchants helping business reduce reputation risk and illegal transactions.• Defined tasks for the project team using Kanban agile methodology for quick prototyping and deployment.• Migrated millions of merchant's data on the AWS RDS MySQL database service reducing operating costs by 30%.• Developed distributed Spark ML pipelines to label raw datasets to achieving better performance over single node.• Developed pipelines to scrape merchant's data from the web into existing dataset on AWS S3 using Scrapy framework.• Developed a predictive model based on word embeddings achieving 95% accuracy on validation dataset.	Oct 2019 - Present
Syngenta Digital Innovation Lab, Champaign, USA <i>Data Science Technology Intern</i> <ul style="list-style-type: none">• Lead the process of designing and implementing the architecture of data science platform on AWS cloud.• Developed a python SDK to connect to private data repositories and create ML environments for quick experimentation.• Delivered an MVP by implementing ML use cases reducing time to implement data science solutions by 40%.• Recommended incorporating AutoML features to the cloud solution saving business more than \$100k in annual costs.	May 2019 – July 2019
School of Information Sciences, UIUC, Champaign, USA <i>Graduate Research Assistant</i> <ul style="list-style-type: none">• Research focused on factors affecting data quality and how scientists make data quality judgment in their field.• Conducted semi-structured user interviews of scientists based on research questions centered on data quality.• Analyze interview data using grounded theory to understand the way data are perceived in the scientific community.• Findings aimed to guide policy decisions in generation and utilization of quality datasets in scientific field.	Jan 2019 - Present
Research IT Technology Services, UIUC, Champaign, USA <i>Data Analyst</i> <ul style="list-style-type: none">• Developed web analytics reports and dashboards using machine data generated from AWS Beanstalk environment.• Explored Splunk integration with external tools for indexing and visualization of search queries and marketing insights.• Devised a strategy to integrate data from different sources for user analytics using Splunk Enterprise.• Used Splunk Machine Learning Toolkit (MLTK) to develop models for anomaly detection, user clustering and web analytics.	Apr 2019 – July 2019
Accenture, Mumbai, India <i>Application Development Associate</i> <ul style="list-style-type: none">• Developed analytics solution to automatically generate business KPIs reducing response time for business decisions.• Increased efficiency of business services by 20% by writing bash scripts to automate data processing and management.• Improved internal tools such as system dashboard and logs management with new features reducing manual processing.• Developed web integrating tool used by project teams to increase cross collaboration increasing productivity by 25%.	Nov 2016 – Jan 2018

ACHIEVEMENTS/CERTIFICATIONS

- Won **Best Intern Award** in “**Best Technological Innovation**” at Research Park Summer Intern awards.
- Awarded “**Six-Sigma Yellow Belt**” certificate for developing analytics solution and increase customer engagement.
- Completed Udacity's ‘**AI Product Manager**’ and ‘**Neural Network Foundations**’ Nanodegree.

PROJECTS

- Indian Scriptures Dataset** - (Numpy, Pandas, Jupyter Notebook, Scrapy) [GitHub](#)
- Created a structured dataset of Indian scriptures by crawling website using Scrapy framework for text mining and NLP.
 - Wrote notebooks to clean and organize raw dataset into individual files for easy training and modeling.

- Dog Breed Classifier** – (NumPy, Keras, Scikit-learn, Matplotlib, OpenCV) [GitHub](#)
- Developed CNN classifier for 133 dog breeds classes using bottleneck features from a pretrained Xception model.
 - Used OpenCV's implementation of Haar feature-based cascade classifier to detect human faces in input images.