

# HARSHIT GUPTA

Champaign, Illinois | [hrgupta@outlook.com](mailto:hrgupta@outlook.com) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

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

## 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

<b>Synchrony Financial, Champaign, USA</b> <i>Data Scientist Intern – Emerging Technology</i> <ul style="list-style-type: none"><li>• Lead the development of an NLP solution to classify suspicious merchants helping business reduce illegal transactions.</li><li>• Migrated millions of merchant's data on the AWS RDS MySQL database service reducing bottlenecks in processing data.</li><li>• Developed Spark ML pipelines to label raw datasets to achieve better performance and reduce time for model evaluation.</li><li>• Compared multiple predictive models on AWS SageMaker with the best model achieving 95% validation dataset accuracy.</li><li>• Developed data pipelines to scrape merchant's data from the web into existing dataset to select better predictive features.</li></ul>	<b>Oct 2019 – May 2020</b>
<b>Syngenta Digital Innovation Lab, Champaign, USA</b> <i>Data Science Technology Intern</i> <ul style="list-style-type: none"><li>• Lead the process of designing and implementing the architecture of data science platform on AWS cloud.</li><li>• Developed python library to connect to private data repositories and create ML environments for quick experimentation.</li><li>• Delivered an MVP by implementing ML use cases reducing time to implement data science solutions by 30%.</li><li>• Recommended incorporating AutoML features to the cloud solution saving business more than \$100k in annual costs.</li></ul>	<b>May 2019 – July 2019</b>
<b>School of Information Sciences, UIUC, Champaign, USA</b> <i>Graduate Research Assistant</i> <ul style="list-style-type: none"><li>• Research focused on factors affecting data quality and how scientists make data quality judgment in their field.</li><li>• Conducted semi-structured user interviews of scientists based on research questions centered on data quality.</li><li>• Analyze interview data using grounded theory to understand the way data are perceived in the scientific community.</li><li>• Findings aimed to guide policy decisions in generation and utilization of quality datasets in scientific field.</li></ul>	<b>Jan 2019 – May 2020</b>
<b>Research IT Technology Services, UIUC, Champaign, USA</b> <i>Data Analyst</i> <ul style="list-style-type: none"><li>• Developed web analytics reports and dashboards using machine data generated from AWS Beanstalk environment.</li><li>• Explored Splunk integration with external tools for indexing and visualization of search queries and marketing insights.</li><li>• Devised a strategy to integrate data from different sources for user analytics using Splunk Enterprise.</li><li>• Used Splunk Machine Learning Toolkit (MLTK) to develop models for anomaly detection, user clustering and web analytics.</li></ul>	<b>Apr 2019 – July 2019</b>
<b>Accenture, Mumbai, India</b> <i>Application Development Associate</i> <ul style="list-style-type: none"><li>• Developed analytics solution to automatically generate business KPIs reducing response time for business decisions.</li><li>• Increased efficiency of business services by 20% by writing bash scripts to automate data processing and management.</li><li>• Improved internal tools such as system dashboard and logs management with new features reducing manual processing.</li><li>• Developed web integrating tool used by project teams to increase cross collaboration increasing productivity by 25%.</li></ul>	<b>Nov 2016 – Jan 2018</b>

## 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

- Adversarial Analysis of Google AudioSet**
  - Implemented multi-level attention model to predict the multi-labels of audio sounds for Google AudioSet.
  - Compared the impact of Siren attack adversarial technique on the performance of different machine learning models.
- Indian Scriptures Dataset**
  - Created a structured dataset of Indian scriptures by crawling web using Scrapy framework for text mining and NLP.
  - Wrote notebooks to clean and organize raw dataset into individual files for efficient training and modeling.
- Dog Breed Classifier**
  - 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.