

# HARSHIT GUPTA

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

## EDUCATION

|  |                 |
|--|-----------------|
| <b>School of Information Sciences, University of Illinois at Urbana-Champaign, USA</b><br>Candidate for Master of Science in Information Management.       | <b>May 2020</b> |
| <b>Thadomal Shahani Engineering College, University of Mumbai, India</b><br>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, Bash, Boto3, Git, Docker
- Cloud Technologies: AWS EC2, S3, VPC, ECS, RDS, Spark MLlib
- Libraries: NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow, Keras
- Data Science Platforms: Jupyter Notebook, RStudio, Splunk Enterprise, Tableau, AWS SageMaker

## PROFESSIONAL EXPERIENCE

|   |                             |
|---|-----------------------------|
| <b>Synchrony Financial, Champaign, USA</b><br><i>Data Scientist Intern – Emerging Technology</i> <ul style="list-style-type: none"><li>• Developing data science solution using NLP to find suspicious merchants helping business reduce fraudulent and illegal transactions worth millions of dollars.</li><li>• Defined tasks for the project team using Kanban agile methodology for quick prototyping and deployment.</li><li>• Migrated millions of merchant's data on the AWS RDS MySQL database service reducing operating costs by 30%.</li><li>• Developed distributed Spark ML pipelines to label raw datasets to achieving better performance over single node.</li><li>• Developed a predictive model for merchant's classification achieving 90% accuracy on validation dataset.</li></ul> | <b>Oct 2019 - Present</b>   |
| <b>Syngenta Digital Innovation Lab, Champaign, USA</b><br><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 platform.</li><li>• Developed python SDK 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 40%.</li><li>• Recommended incorporating AutoML features to the cloud solution saving business approximately \$100k in annual costs.</li><li>• Won <b>Best Intern Award</b> in "<b>Best Technological Innovation</b>" for the platform at Research Park Summer Intern awards.</li></ul>               | <b>May 2019 – July 2019</b> |
| <b>Research IT Technology Services, UIUC, Champaign, USA</b><br><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>School of Information Sciences, UIUC, Champaign, USA</b><br><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 - Present</b>   |
| <b>Accenture, Mumbai, India</b><br><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><li>• Awarded "<b>Six-Sigma Yellow Belt</b>" certificate for developing analytics solution and increase customer engagement.</li></ul>                           | <b>Nov 2016 – Jan 2018</b>  |

## PROJECTS

|  |
|--|
| <b>Adversarial Analysis of Google AudioSet - (NumPy, Pandas, TensorFlow, Scikit-learn, Matplotlib)</b> <ul style="list-style-type: none"><li>• Implemented multi-level attention model proposed for weakly labeled datasets such as AudioSet for better accuracy.</li><li>• Compared impact of Siren attack adversarial technique on the performance of different machine learning models.</li></ul> |
| <b>Dog Breed Classifier – (NumPy, Keras, Scikit-learn, Matplotlib, OpenCV)</b> <ul style="list-style-type: none"><li>• Developed CNN classifier for 133 dog breeds classes using bottleneck features from a pretrained Xception model.</li><li>• Used OpenCV's implementation of Haar feature-based cascade classifier to detect human faces in input images.</li></ul>                              |
| <b>TV Scripts Generator – (NumPy, TensorFlow, Scikit-learn)</b> <ul style="list-style-type: none"><li>• Developed a Recurrent Neural Network to generate The Simpsons TV scripts using word embeddings.</li><li>• Implemented data preprocessing techniques and created input batches for accurate sentence generation.</li></ul>  |