HARSHIT GUPTA

San Jose, California | hrgupta@outlook.com | 🔗 LinkedIn | ⌂ GitHub

EDUCATION

School of Information Sciences, University of Illinois at Urbana-Champaign, USA   **May 2020**
Candidate for Master of Science in Information Management.

Thadomal Shahani Engineering College, University of Mumbai, India   **May 2016**
Graduated as Bachelor of Engineering in Computer Engineering with First Class.

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   **Oct 2019 – May 2020**
Data Scientist Intern – Emerging Technology
- Lead the development of an NLP solution to classify suspicious merchants helping business reduce illegal transactions.
- Defined and managed 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 bottlenecks in processing data.
- Developed Spark ML pipelines to label raw datasets to achieve better performance and reduce time for model evaluation.
- Compared multiple SVM models on AWS SageMaker with the best model achieving 95% validation dataset accuracy.
- Developed data pipelines to scrape merchant's data from the web into existing dataset to select better predictive features.

Syngenta Digital Innovation Lab, Champaign, USA   **May 2019 – July 2019**
Data Science Technology Intern
- Lead the process of designing and implementing the architecture of data science platform on AWS cloud.
- Developed python library 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 30%.
- Recommended incorporating AutoML features to the cloud solution saving business more than $100k in annual costs.

School of Information Sciences, UIUC, Champaign, USA   **Jan 2019 – Present**
Graduate Research Assistant
- 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.

Research IT Technology Services, UIUC, Champaign, USA   **Apr 2019 – July 2019**
Data Analyst
- 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.

Accenture, Mumbai, India   **Nov 2016 – Jan 2018**
Application Development Associate
- 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%.

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 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.