

# Navita Jain

navita.jain01@gmail.com | 518-369-3201 | Dublin, CA, 94568 | LinkedIn | GitHub

## Experience

---

**Data Analyst**, *Girl Scouts of Northeastern New York* Feb 2017- Apr 2019

- Applied machine learning, natural language processing and statistical analysis techniques to generate insights that facilitated in data-driven strategic business decisions.
- Performed exploratory data analysis(EDA), trend and pattern analysis, anomaly detection and data visualization to better understand the data.
- Performed feature engineering, feature selection and feature importance using ensemble methods like Random Forest and Gradient Boosting.
- Responsible for end-to-end data analysis including data preprocessing, manipulation and aggregation on structured and unstructured data using Python. Presented findings to directors and executive staff.
- Responsible for developing ETL(Extract, Transform and Load) processes and large scale data migration from a legacy system to cloud based system.
- Automated data mapping from multiple systems(Donor, Member etc) into a single source file.
- Developed a Python Application using TkInter and PyInstaller as a support tool for data migration.
- Wrote SQL queries for reports and dashboards
- Developed modules in C# for a cloud-based system called iMIS to act as an import utility and integrated solutions for data manipulation and mapping to enable bulk uploads for the data from Salesforce system into iMIS.

**Graduate Research Assistant**, *Data Mining lab, University at Albany* Apr 2015 - Jan 2017

- Developed efficient statistical and data mining techniques in Python for community detection in social networks.
- Implemented two state burst detection model on time series data for graph network using dynamic programming.

**System Engineer**, *Infosys Limited* Mar 2010 - Sep 2012

- **Oracle DBA** - Responsible for database administrative activities for a large and complex Oracle 11g and 9i database system. Developed and modified SQL scripts for changing business requirements.
- **Salesforce Developer** - Implemented, administered and maintained client's Salesforce business application. Created Custom Objects, Custom fields, Custom Tabs, Reports and other components as per requirements.

## Education

---

**Master of Science, Computer Science**, *State University of New York, Albany* Dec 2016

**Online Course: The Data Scientist's Toolbox**, *Coursera, by Johns Hopkins University* Jun 2014

**Bachelor of Technology, Information Technology**, *Rishiraj Institute of Technology* Sep 2009

## Skills

---

**Languages** | Python, Java, SQL, JavaScript, HTML

**Machine Learning** | Clustering, Classification, Regressions, Feature Selection.

**Analytics Packages** | Pandas, NumPy, SciPy, Matplotlib, scikit-learn, TkInter

**Tools, Services** | GitHub, Amazon Web Service(AWS), Spark, Hadoop, MySQL, Jupyter Notebook, Apache Lucene

## Projects

- 
- **Master's Thesis - Synchronized Communities in large scale graph networks**, Performed synchronised behaviour analysis using statistical and hypothesis testing on large social networks
  - **Natural Language Processing - Influence detection using Politeness and Likeability**, Applied Support Vector machine on Twitter dataset
  - **Software Engineering - Online Voting Application: BallotBox**, Build Voting module using MVC model, Developed MySQL database on AWS RDS and hosted the web application on AWS Elastic Beanstalk Tomcat server
  - **Information Retrieval - Document Retrieval for Explicit Temporal Queries**, Proposed an explicit time sensitive TF-IDF(Term Frequency - Inverse Document Frequency) variant for Lucene search engine
  - **Data Mining - Discovering Social Circles**, Applied K-means clustering and Jaccard Similarity for categorization of nodes in social clusters, used Balanced Error rate and F1 score for evaluation