

ELLISA KHOJA

LinkedIn: <https://www.linkedin.com/in/ellisa-khoja/>
Github: <https://github.com/ellisa1419>

Email : ellisa.khoja@gmail.com
Mobile : 518-423-7196

EDUCATION

- **University at Albany, SUNY** Albany, NY
Master of Science in Computer Science; GPA: 3.8 Aug 2017 - Dec 2018
- **Kalol Institute of Technology and Research Center** Gujarat, India
Bachelor of Engineering in Computer Engineering; GPA: 3.4 Aug. 2010 – July. 2014

PROGRAMMING SKILLS

- **Technologies:** Python, Java, AWS, Kafka, Docker, Kubernetes, REST/Web Services, Spark, Angular
- **Databases & Tools:** Pycharm, Git, MySQL, Postgresql, Oracle , NoSQL, Cassandra, MongoDB, InfluxDB
- **Achievements:** 5th place in Leetcode weekly contest, Make A Difference Award - Delivery, Cignex Datamatics

EXPERIENCE

- **University at Albany - Research Assistant** Albany, NY
Sept 2017 - Present
 - **PANACEA:** Working with an interdisciplinary team across multiple sites with the goal of identifying the Social Engineering attacks in emails for D.A.R.P.A ASED challenge. Built the stylistic classifier for enron email dataset for authorship verification using Python Email Parser. Predicted DialogActs for chat/email conversations using lexical-syntactical features like N-grams, Stanford POS, Named Entity Tagger and NLTK library.
 - **SimOn:** Applied Data Mining and Data preprocessing techniques on dataset of about 10M records using Python Scripts in Github network. Simulated events for the top 1 million repositories by implementing Markov chains based on sequence of events. Clustered the repositories based on their activeness and events in GitHub using K-Means for a given timeframe.
- **Edwards Lifesciences - Software Engineer Intern** Irvine, CA
May 2018 - August 2018
 - **Critical Care - Algorithms:** Collected data from unstructured format, pre-processed and imported the into PostgreSQL using Python SQLAlchemy. Parallelized the writes to the improve the speed and wrote API to read/write the queries using Python. Explored NoSQL databases for timeseries data. Imported data into Apache Cassandra database using Java library of SSTables and Apache Spark on ec2 instance. Compared performance of Cassandra, InfluxDB, TimeSeriesDB and databricks for speed; tuned database parameters to improve results by over 20 percent. Managed the servers on AWS Ec2 instance, RDS and S3 for deployment.
- **CIGNEX Datamatics - Software Engineer** Ahmedabad, India
July 2014 - July 2017
 - **Hotel Operations Management:** Worked on Hotel Operation project to migrate the project using Java and Spring framework. Worked on Front side and Server side to implement business logic using MVC design and performed database cleaning. Created and consumed Web Services for restful Calls, and worked JWT tokens for authorization. Configured hudson build to deploy application on production and followed Agile workflows.
 - **Franchise Management Platform:** Developed application in Java and Liferay to manage Franchise operations such as create franchise hierarchy for organizations and subdomains, CMS to share content amongst Franchises and assigning roles to users, created messaging portal for posting, announcement and generate notifications, created revenue reports using Jasper and worked on inspection Modules. Created rest Services to expose data. Worked on both Front end and Backend and Database design . Used Java, Liferay 6.x, JQuery, Angular, MySQL, Rest API

PROJECTS

- **Housing Prediction using Regression:** Predicted housing price of Kaggle data set, using regression techniques, preprocessed and performed data analysis and feature selection. ML : Logistic Regression.
- **Movie Recommendation System:** Implemented Movie Recommendation system to suggest movie based on past movies and user profile using Association Rule Mining and Logistic Regression.
- **Query Expansion Using Wordnet:** Query Expansion using Wordnet to add synonyms for each token in query to improve the precision and recall using lucene search engine in Information Retrieval. Languages: Python, Lucene