# Ray Mohabir

(347) 924-2797 • raymohabir1@gmail.com • LinkedIn • New York, NY

## EDUCATION

**New York University,** *New York, NY* — MAY 2021
Degree: Master of Science in Data Science — GPA: 3.8 / 4.0
Relevant Coursework: Natural Language Processing, Big Data, Machine Learning, Probability and Statistics

**New York University,** *New York, NY* — MAY 2018
Degree: Bachelor of Science in Computer Engineering — GPA: 3.6 / 4.0
Relevant Coursework: Multivariate Calculus, Design and Analysis of Algorithms, Embedded Systems

## PROFESSIONAL EXPERIENCE

**GeneCentrix Inc** — Sep 2020 – Dec 2020
Data Scientist (Capstone) — New York, NY

- Improved data gathering techniques for biomedical texts consisting of drug-target interactions for small molecule drugs
- Developed a classification model to locate named entities from the collected biomedical texts using a biomedical language representation model
- Enhanced the quality of the existing NLP classification model from 84% to 96% using Python, NumPy, SciSpaCy
- Implemented an end-to-end framework to handle drug upscaling using Fuzzy String Matching and automated the labeling process of 10,000+ items
- Refined an existing SciSpaCy model to identify drug-target interactions using Named Entity Recognition

**Advanced skills:** Python, NLP, Pandas, NumPy, SciSpaCy, Machine Learning, Data Analytics

**NYS Department of Financial Services** — Jan 2018 – May 2018
Capital Markets Analyst — New York, NY

- Spearheaded the creation of a relational database (1,500+ records) using MS Access, SQL, and Excel to store check-cashing data suited to the specifications of the Deputy Director
- Researched and compiled subprime auto loan securitization data from 20+ sources to support the Capital Market's division in communicating its objective to the public
- Co-authored the final report detailing statistics regarding subprime auto loans for 1,500+ zip codes in NYS
- Analyzed check-cashing financial statements for profit and loss analysis using MS Excel

**Advanced skills:** SQL, MS Access, MS Excel, Data Analytics

## PROJECTS

**Analysis of Serendipity in Recommender Systems** *(Skills: Python, Pandas, NumPy)*

- Performed exploratory data analysis and applied box-cox transformation and outlier trimming for data normalization
- Implemented three serendipitous algorithms and a baseline to analyze diversity, serendipity, and novelty in recommendation systems
- Identified a trade-off between serendipity in recommender systems and accuracy in matrix factorization models

**Fine-tuning RoBERTa for Content Moderation** *(Skills: Python, NLP, RNN, Data Analysis, Statistics)*

- Implemented an automatic comment-flagging model specific to the rules of a given subreddit using a RoBERTa model
- Created a recurrent neural network model as a baseline algorithm to compare the identification of problematic content
- Achieved a F1 score of 85% and improved the baseline model's accuracy from 74% to 77%

**Movie Recommender System** *(Skills: Python, Neural Networks, Recommender System)*

- Spearheaded a team to vigorously evaluate algorithms useful for recommender systems and crafted a business proposal for the recommender systems implemented
- Tested collaborative matrix factorization, SVD, and neural networks in Python to develop recommender system models
- Published the final business report detailing the importance of the recommender system and its performance

**Circuit Component Classification** *(Skills: Python, Keras, CNN)*

- Researched algorithms for multi-object classification on 5 circuit components. Achieved an overall accuracy of 93%
- Built the framework for the convolutional neural network to identify key features and implemented using Python

## TECH SKILLS

| | |
|---|---|
| Programming: | Python (NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn), SQL, C++ |
| Data Science: | NLP, Big Data, Machine Learning Models, Probabilistic Models |
| Machine Learning: | Supervised (Classification, Regression), Unsupervised (PCA, SVD) |
| Data Engineering: | MySQL, MapReduce, PySpark, Hadoop, PyTorch, Keras |
| Statistical Inference: | A/B Testing, Probability, Statistics, Correlation, Regression |