# NAMAN SHARMA

## DATA SCIENTIST

## CONTACT

- ✉ namansharma.0992@gmail.com
- 📞 (551) 229-5952
- 🌐 New Jersey / New York
- in linkedin.com/in/2namansharma
- ⌂ github.com/namansharma0992

## EDUCATION

2018 New Jersey Institute of Technology
**Master of Computer Science**

2014 Amity University [Gurgaon, India]
**Bachelor of Technology in Computer Science**

## TECHNICAL SKILLS

- Machine Learning
- Software Engineering
- Advanced Analytics
- Statistics and Probability
- Data Visualization
- Python (numpy, pandas, scipy, scikit-learn, xml, json, beautifulsoup, matplotlib,), R (ggplot2, dplyr), SQL
- Jupyter Notebook, RStudio, Git/Github, KNIME

## MACHINE LEARNING ALGORITHMS

- Gaussian Naïve Bayes, Support Vector Machine, Decision Trees, Gradient Boosting, Random Forest, Linear Regression, K-mean Clustering, Lasso, K-Nearest Neighbors
- Feature Engineering, Text Learning, Principle Component Analysis, Cross Validation, Grid Search, Ransom Search, Model Selection, Model Tuning.

## EXPERIENCE

**Data Science Engineer |** Nova IQ **|** March 2019 - Dec. 2019
- Developed Web content monitoring system to scrape real-time market insights using selenium and python resulting in saving of $400,000 per annum.
- Coded 12 specific scrapers for specific market data and 2 generic scrapers for the remaining web.
- Coded a text-summarizer using python and its library nltk to summarize the content that was scraped.
- Worked on the development of Application Tracking System (ATS) demo using Pytorch Library and Recurrent neural network for client onboarding

**Data Analyst Intern** | BCT Partners | Feb 2018 – May 2018
- Created end-to-end data science workflow by wrangling 2+ GB of data from multiple sources, cleansing and harmonizing the data, training predictive models using KNIME analytics platform and Python.
- Engineered 200+ new features before completing feature selection and model tuning for random forest model.
- Optimized software to reduce runtime from 7.5 hrs to 20 mins for 1gb dataset.
- Built interactive interface to allow non-technical users to easily leverage workflow with new datasets.

**Data Science Projects**
**Identifying Financial Fraud at Enron**
- Built model to identify Enron insiders who committed fraud by analyzing financial records from 158 former employees using Gaussian Naïve Bayes classifier in Python(sklearn).
- Tested multiple models, including tree-based approaches and nearest-neighbor models, to identify the ideal approach
- Optimized model by identifying and removing specious outliers, engineering critical new features, standardizing data, removing nonpredictive features, and using grid search with cross-validation to tune hyperparameters
- Achieved over 50% precision and 38% recall identifying fraudulent individuals.

**Statistical Testing of The Stroop Effect**
- Analyzed the Stroop Effect using descriptive statistics and statistical tests in Python (scipy). Performed dependent t-test with a confidence interval of 99% and ultimately rejected null hypothesis based on differences found in user response times to congruent and incongruent tasks.

**OpenStreetMap New York City Data Assessment**
- Assessed the validity, accuracy, completeness, consistency and uniformity of over 2.6GB of NYC map data from Open Street Maps using data wrangling techniques in Python and SQL. Modified and removed unqualified data.