Kumar Kovid

(201) 969-6646 | 85 Central Avenue, Jersey City, NJ, 07306 | kkovid@stevens.edu | Github | LinkedIn

EDUCATION

Stevens Institute of Technology **Aug 2018 - May 2020**
Master of Science in Information Systems | GPA 3.94/4 **New Jersey, United States**
Coursework: Web Mining, Data & Knowledge Management, Marketing Analytics, Intro to Bloomberg

Government College of Engineering and Technology **Sep 2013 - Jul 2017**
Bachelor of Engineering in Electrical Engineering **Jammu, India**

SKILLS

Programming Languages & Web Frameworks: Python, SAS, R | HTML5, CSS3, Flask
Tools: Advanced Excel, Tableau, Power BI, Hadoop, Spark, Google Analytics, Informatica, MS Project, Git, Docker, AWS
Databases: My SQL, SQLite, Oracle, Microsoft SQL server, Postgres SQL, DB2
Python Libraries: Pandas, Numpy, Scikit Learn, TensorFlow, re, Nltk, Seaborn, Matplotlib, Selenium, beautifulsoup4, pickle
Algorithms: Naive Bayes, Logistic Regression, Linear Regression, Random forest, Gradient boost, SVC, K-means, LDA, KNN

WORK EXPERIENCE

NYC Department of Finance **Sep 2019 - Dec 2019**
Data Scientist (Fall Intern) **New Jersey, United States**

- Analyzed property, fire, water & permits data using numpy & pandas, across all 5 boroughs of NYC with the Director of Property Analytics Group, identifying almost 4K new property tax evasion cases
- Leveraged regression, classification & statistical framework models on NYC's property, permits & complaints data to discover 1K illegal constructions from 1.1 million properties, reducing assessment cost by 20%
- Investigated these extracted illegal construction & tax evasion cases for risk & revenue, using additional parameters in pandas & numpy to scale down to top priority cases for immediate attention
- Designed reporting requirements & expedited ad-hoc, weekly & monthly quantitative analysis reporting on KPI's using SQL queries, Excel VBA by drawing Pivot Tables & Charts, to reduce weekly reporting time by 8 hours

Data Scientist (Summer Intern) **Jun 2019 - Aug 2019**

- Developed a KNN model in Python to find market value of non-valued buildings based on similarity with valued-buildings
- Created SAS & Python programs to extract historical building statistics using NLP & RegEx, from Assessor's textual notes for 50K buildings & compared data points with 3-D mapped view to discover data irregularities
- Developed a star-schema relational database after integrating data from MS Excel, MS Access & SAS, reducing formatting & mapping errors by 40%

Tata Consultancy Services (TCS), India **Mar 2018 - Aug 2018**
Assistant Systems Engineer **Delhi, India**

- Analyzed employee data to identify reasons of high attrition, leveraging SVM & Random Forest & correlation models
- Developed & optimized Informatica workflows successfully & generated pivot reporting using advanced Excel & Visio
- Spearheaded 10% improvement for on-time deliverables by supervising communication between IT & business teams
- Collaborated with cross-functional teams to update & maintain 2 internal databases for data analysis & research

ACADEMIC PROJECTS

Movie Review Sentiment Analysis, Personal Project| Python | Project **Sep 2019 - Dec 2019**

- Extracted movie reviews from rotten-tomatoes using Selenium & manually classified them as positive or negative
- Implemented normalization, stemming, lemmatization & kept only adverbs, adjectives & verbs as allowed words
- Predicted polarity & sentiment of unlearned review as an aggregation of 7 classifiers including Logistic, MNB & SVC

Anti Human-Trafficking System | UN Office of Drugs & IBM Hackathon | SQL | Python | Project **Mar 2019 - Jun 2019**

- Investigated relation between NYPD crime data & malicious Ads location, time & phone number of posting
- Established a knowledge graph to depict 100 crime & Ad relations with node4j, winning invitation to IBM Horizons

E-mail Spam Detector | Stevens Institute of Technology | Python | Project **Feb 2019 - Mar 2019**

- Trained Random Forest & Gradient Boost model using labeled email data to classify un-labeled emails as spam or ham
- Engineered length & punctuation% from email as features & optimized model parameters using GridSearchCV
- Achieved an overall accuracy of 97.6 %**,** recall rate of 84.5 %, prediction time of 0.115 seconds & precision of 90.8%

LEADERSHIP ACTIVITIES

GCET, India: Partnered with a 5-member team to organize robotics workshop for IIT-Kanpur **Dec 2019 - Dec 2019**
Stevens, USA: Organized 2 badminton competitions with all 5 disciplines & more than 50 competitors **Apr 2019 - Nov 2019**