Neha Gupta
Software Engineer, Data

✉ neha.atul20@gmail.com
📞 984-289-5129
📍 San Francisco, CA

in nehaguptas
⭘ ngupta8

Education

California State University, Los Angeles GPA: 3.97

MS Information Systems 2019
Received College of Business and Economics Honors
Award for 4.0 GPA

Skills

Programming: Python, SQL, NoSQL
Machine Learning: NLP, Classification, Regression, Recommendation system
Tools/Technologies: Data Warehousing, ETL, AWS, Spark, Hadoop, Excel, Hive, Databricks, Pandas, NumPy, Scikitlearn, Beautiful Soup, Selenium, Stats models, CI/CD, SDLC, A/B Testing
Visualization: Tableau, Flask, Seaborn, Matplotlib, Bootstrap

Summary

Excellent team player with an analytical mindset, I am seeking a challenging opportunity where I can apply my knowledge and learning from my master's degree about machine learning, data analysis and programming skills and techniques. I would like to expand my experience in building data pipelines to migrate and analyze data. I have experience in data analysis, building reports and presenting them to respective stakeholders using visualization techniques.

Experience

Data Scientist, Metis   San Francisco, CA   Jan. 2019 to Current.
Personalized Restaurant Recommender(AWS, Python, Recommender System, Flask, BootStrap):
- Built a recommender system using Yelp data in order to provide personalized restaurant experience. The project includes Content-based, Collaborative and Hybrid approach
- Used KNN and SVD model to check the performance on ratings. Application created with help of Flask, Bootstrap
- **Github link:** github.com/ngupta8/Yelp-Recommender

Online Bidding, Human or Bot(AWS, PySpark, SparkSQL , Matplotlib, Python, Flask, BootStrap):
- Built a service that predicts and identify Ad Fraud by isolating human bids from computer-generated bids
- Prepared the Random Forest Model from available training data. Used PySpark, SparkSQL for feature engineering
- **Github link:** github.com/ngupta8/Online-Bidding-Human-or-Bot

Graduate Student, California State University Los Angeles   Los Angeles, CA   Aug. 2017 to May 2019
Crime Pattern Analysis(Hadoop, Hive, Sql, Tableau):
- Built a system to analyze crime data in metropolitan cities in USA to identify and interpret patterns.
- Built basic reverse geocoding with Python UDF in Pig to get zipcodes from raw data with latitude and longitude Used Hadoop HDFS to store data and Hive queries to identify most crime-prone areas in metropolitan cities regarding day, time and zipcodes Used Tableau for visualization of refined data
- **Github link:** github.com/ngupta8/Crime-Pattern-Analysis

AdTracking Fraud Detection(AWS, Databricks, PySpark, Python):
- Built a system to analyze user access patterns to predict whether an app will be downloaded after clicking an ad or not Used PySpark, Databricks for analysis and predicting the fraud
- Used data from a Kaggle competition: "TalkingData AdTracking Fraud Detection Challenge"
- **Github link:** github.com/ngupta8/Predict-AD-click-fraud

Data Analyst Intern, IT Strategists   Los Angeles, CA   June 2018 to July 2018
- Worked on an ETL migration project to translate SQL based ETL pipeline to Python-based ETL pipelines. The new pipeline is more maintainable and observable making it reliable and predictable
- Analyzed BI reports and the SQL queries to understand the complexity of the system based on rules defined by the business
- This analysis helped customer to streamline their data warehouse and choose the right tool to be more effective and efficient

Data Analyst, One97 Communication Pvt.   India   May 2011 to July 2015
- Gathered applications requirements from clients and made the flow diagram on Microsoft Visio
- Written SQL queries and scripts to extract and analyze data, and to create automated self-service reports
- Analyzed and interpreted data to understand usage and behavior trends of subscribers
- Building KPI reports and dashboard. Also, fixing issues in existing reports.
- Meet the application development deadlines with highest quality.