Dhrumil Hiten Vora

Email: dhrumilv93@gmail.com

Phone: +1 (540) 405-7574

Jersey City, NJ - 07306

Website | LinkedIn | GitHub

Summary

Pursuing M.S in Computer Science expected graduation in *May '18* specializing in **Databases and Data Mining** with greater than 1-year of work experience in **Development** and **maintaining DB** with interest in finding solutions to business problems by interpreting and analyzing data and seeking to leverage data analytical skills to improve corporate performance as a data analyst.

Projects

Finding Missing Cards | Hadoop, AWS EC2, MR GitHub   **Intro to Big Data in MS**   *Mar '18*

- Main goal of this program to get used to the Big Data tools such as Hadoop, Amazon Web Services EC2 instances and interaction between them and a basic understanding of Map Reduce programs such as workflow, MR Driver, Mapper, Reducer to find missing cards from a deck of 52 cards.

Finding highest relative frequency of words | Hadoop, AWS EC2, S3, MR   **Intro to Big Data in MS**   *Mar '18*

- Developed a challenging MapReduce Program which analyses 100K wiki articles and provides the result of relative frequency of every word with every other word in the articles. The result shows top 100 words with highest relative frequencies.

News Article Parser | Python, Beautiful Soup, Requests and CSV – GitHub   *Sept '17*

- Built a program which can parse all the articles from the specific number of pages from the archive of website www.dailymail.co.uk. For fetching the html web pages library used is request and for parsing the HTML pages, XML, JSON responses from the API calls the library used is Beautiful Soup
- Extracting the Article Title, Author, Published and Updated date and the Content of the article and converted the extracted data into Excel CSV format.

Supervised data mining | Java, Weka tool   **Data Mining in MS**   *Oct '17*

- Analyzed the performance comparison using the Weka tool between two learning algorithms Decision Trees and Naïve Bayes for a dataset that consisted the classification of phishing websites.
- The outcome was that the Decision Trees algorithm (overall 97.26%) was ~5% more accurate than the Naïve Bayes algorithm (overall 92.98%) for the specific dataset of phishing websites.

Guestbook Webapp | PL/SQL, Oracle 10g, Python and Django - GitHub   **Advance Database in MS**   *Sept '17*

- The objective was to build a guestbook that lets users post messages to a public message board using HTML and Django.
- The challenges faced in this project was the process of developing a connection between the Django project and the Oracle database.

Classification of Satellite Images | MATLAB – GitHub   *Dec '16*

- Designed the project to classify the satellite images based on histogram technique in four categories (costal, forest, desert, metro) with the use of reference images of all four categories and with the accuracy of ~92%.

Work Experience

Synergics Solutions – Data Analyst and Product Management Intern   *Jan – Jul '16*

- Automated operational process (data monitoring, reports for clients etc.) using Python and SQL.
- Troubleshooting data inaccuracy and breakdown in operational process.
- Built data visualizations dashboards and analyze data of major clients using Tableau.

ITSource Technologies – Data Analyst Intern   *Jun – Dec '15*

- Used statistical techniques for hypothesis testing to validate data and interpretations.
- Leveraged MS Excel skills (Pivot Tables, V-Lookups, etc) to provide analytics to stakeholders also using Tableau dashboards.
- Built a survey form for new users which could ask the user to select from many type of cuisines to generate the user profile for recommendations with the accuracy of 75%.

JAE Pharmacy – Data Analyst Intern   *May – Aug '14*

- Communicated with the clients to gather the analytics requirement and ensure the successful delivery of the projects.
- Performed Descriptive Analytics of the clients data and presented in the form of Tableau visualizations.

Technical Proficiency

- **Development Skills:** Python (NumPy, Pandas, Beautiful Soup, SQL Alchemy, scikit-learn), Java, Map Reduce, C++, C#, MATLAB.
- **Web Dev skills**: HTML5, CSS3, JavaScript, Bootstrap4.
- **Databases:** MySQL, Oracle 10g, 11g, MS SQL.
- **Operating Systems:** Linux, Windows, MacOS X, Ubuntu.
- **Software Skills:** Amazon Web Services (EC2, S3, EMR), GitHub, Hadoop, Tableau, Alteryx.

Education

M.S in Computer Science - New Jersey Institute of Technology, Newark, NJ**.**   *May '18*

B. E in Computer Engineering - Thakur College of Engineering and Technology, Mumbai, India.   *Jun '15*

Certifications: Stanford Online – Databases DB5 SQL.   *Aug '16*

Intro to Hadoop and MapReduce - Udacity   *Feb'18*

Python for Data Science Essential Training- Lynda.com   *Apr'18*

Machine Learning by Andrew Ng -Coursera   *In Progress*

Affiliations

- Deep CS (official graduate student community at NJIT) Tutor for Operating System Design and Internet and Higher layer protocols for 1 year.
- Teaching Assistant for CS101 course Computer Programming where I've Graded the assignments and the in-class exercise which had MATLAB programs.
- **Earned 2nd rank** in Grid Warrior (Cassiopeian Wars) with a team at IIT Bombay Techfest.
- Ranked under **top 1000** all over India in CodeVita a coding competition held by Tata Consultancy Services.