

Aditya Mahajan

(972)-855-7961: adityamahajan1392@gmail.com

[linkedin.com/in/adityamahajan2](https://www.linkedin.com/in/adityamahajan2)

github.com/Murdocc007

EDUCATION

M.S., Computer Science, Data Science

The University of Texas at Dallas

May 2017

GPA **3.66**

B.E., Computer Science

Panjab University, Chandigarh, India

June 2014

CGPA **8.3/10**

WORK EXPERIENCE

Directi, Mumbai, India

October 2014-June 2015

Software Development Engineer

- Developed a fuzzy search for college/institution lookup for registering candidates on the website
- Programmed custom judges for problems in order to check solution to coding problems
- Wrote CRON jobs to daily extract the rank/popularity on websites like Alexa, LinkedIn, Facebook and Twitter

Alert Enterprise Inc., Chandigarh, India

January 2014 - September 2014

Software Development Engineer

- Wrote the backend queries for the incident monitoring dashboard
- Resolved the application defects to facilitate stable releases
- Redesigned legacy backend to integrate with the new badge designer UI

TECHNICAL SKILLS

Programming Languages and Tools: C/C++, JAVA, Python, JavaScript, Django

Databases: Oracle, MySQL, SQL server

Operating Systems: Linux, Windows, MAC OS

Data Science Languages and Tools: R, Scikit Learn, Pandas, Spark, Hadoop, Mapreduce, Hive, Ipython, Pig

PROJECTS

Integer Sequence Learning

- Used kaggle dataset to predict the next number in the integer sequence
- Use linear regression to predict the result
- Ranked 30 out of 200 participants in the competition

Credit Score Analysis

- Used the kaggle dataset to analyze the possibility of defaulting of a person in the next two years
- Used Apache Spark, Hive and ML pipeline to model the data. Made three different input models, using logistic regression, gradient boosting and random forest trees.

Character Recognition

- Given a matrix with pixel intensities, made a model to recognize a character.
- Used Support Vector Machine to model the data
- Extracted best features using Principal Component Analysis

Log Analysis

- Analyzed NASA's one month of HTTP logs using PySpark
- Extracted information like most frequent hosts, top 10 endpoints and the number of daily requests

ACHIEVEMENTS

- Google Code Jam- Reached the second round and secured a rank of 883 out of 10000 participants.
- Ranked 245 out of 4000 participants in the Santander prediction challenge on Kaggle.