Aditya Mahajan

(972)-855-7961: adityamahajan1392@gmail.com
linkedin.com/in/adityamahajan2
github.com/Murdocc007

EDUCATION

M.S., Computer Science, *Data Science* May 2017
The University of Texas at Dallas GPA **3.66**

B.E., Computer Science June 2014
Panjab University, Chandigarh, India CGPA **8.3/10**

WORK EXPERIENCE

CodeChef, Mumbai, India October 2014-June 2015
Software Development Engineer

- Developed a fuzzy search for 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 - September2014
Software Development Engineer

- Wrote the backend queries for the incident monitoring dashboard
- Redesigned legacy backend to integrate with the new badge designer UI

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.

TECHNICAL SKILLS

Programming Languages and Tools: Python,Java, JavaScript,Django,Scala
Databases: MySql, SQL server
Data Science Languages and Tools: R, Scikit learn,Spark,Hadoop,Mapreduce, Hive,Pig,MLPipeline,PySpark

PROJECTS

Cuckoo Hashing

- Using two hash functions increased the load factor to 0.80 as compared to Java's Hash map load factor of 0.75

Online Fiddler

- Made a website to fiddle with javascript,HTML and CSS using Django
- Added a functionality to save existing templates and loading the earlier ones.
- Users can make their templates secret or can make it public for everyone to else to see.

Integer Sequence Learning

- Used kaggle dataset to predict the next number in the integer sequence,using linear regression
- 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.

Publications

- Feature selection using CUR matrix decompositions.(unpublished)