Arpith Thomas Varghese

EDUCATION

State University of New York, Stony Brook
Masters in Computer Science
August 2016 - December 2017
GPA: 3.83

National Institute of Technology, Calicut India
B.Tech in Computer Science and Engineering
July 2009 - June 2013

SKILLS

Languages:
C, C++, Java, Python, SQL, Scala
Tools:
git, maven, spark, memcache,membase, redis, hadoop, scalding

COURSEWORK

Operating Systems
Analysis of Algorithm
Machine Learning
Natural Language Processing
Data Science Fundamentals
Artificial Intelligence
System Security
Big Data Analytics

EXPERIENCE

Twitter May 2017 – August 2017
Software Engineer San Francisco, CA

- Core Data & Metrics

PayPal May 2017 – August 2017
Software Engineering Intern San Jose, CA

- Built a tool in Java which given a git commit figures out which methods were changed as part of the commit and executes only test cases that directly exercise the deduced methods.

Zynga September 2014 – May 2016
Software Engineer Bangalore, IN

- Worked on the complete life cycle of a mobile game from conception to shipment on both iOS and Android stores.
- Developing core gameplay features including a new game mode, in-game achievements and social features like friends and messaging in C++.

Oracle August 2013 – August 2014
Member of Technical Staff Bangalore, IN

PROJECTS

Detecting and rectifying signs of nervousness in public speaking videos

- Used speech recognition library Kaldi to train a LSTM based model to transcribe audio to text
- Detected dis-fluencies in the transcribed text by using Max-Margin Markov Networks and removed it from the original audio.

Per Process System Call in Linux :

- Created a linux kernel based system to support dynamic addition of system calls using modules.
- Added support to customize and block system calls for each process using system call vectors

Stackable Filesystem in Linux :

- Created a stackable filesystem called trfs, build on top of wrapfs, to support tracing of system call operations to a log file in a consistent manner.
- Created user level program to replay filesystem operations written in the log file and verify log file integrity.

Predict a match between two online dating profiles

- Ranked 3rd in the in-class Kaggle Competition to predict possibility of a match between two people, based on their features. Used Polynomial Regression, PCA and engineered features to achieve AUC of 0.67706.

Extracting Key-phrases and Relations from Scientific Publications

- The goal is identify all the key-phrases in a scientific document and to classify them into one of the three categories - Process, Task, Material.
- Key-phrase Extraction was done using Conditional Random Field.
- Trained a char CNN and BLSTM model in Keras to classify them.

Malicious Browser Extension

- Created a browser extension for Google Chrome to steal user information like browsing history, cookies, username and password without user knowledge.
- Added additional functionality to manipulate DOM content of any page to inject custom DOM and run custom javascript code.