225 East 95th St
New York NY – 10128

Pawandeep Singh

singh@alumni.emory.edu
(678)-557-0528

Education

Emory University — Atlanta, GA
M.S. In Computer Science (Machine Learning Concentration) — Aug 2013-May 2015

- Completed courses: Data Mining, Machine Learning, Natural Language Processing, Numerical Analysis, Applied Linear Models, Big Data Analytics.

Guru Tegh Bahadur Institute of Technology — New Delhi, India
Bachelor of Technology in Computer Science — Aug 2007-May 2011

- Completed courses: Artificial Intelligence, Object Oriented Programming Concepts, Data Structures, DBMS

Work Experience

Emory University — Atlanta, GA
Research Assistant — June 2014-Nov 2015

- Performed complexity analysis of Personnel Law data over a period of time, measuring entropy and Flesch score. Measured relative complexity across different countries and relative increase in rate of complexity.
- Performed data wrangling and, munging to obtain data from different countries government sites. Further performed data cleaning and data audit to ensure quality.
- Designed the experiment and validated it at individual stages for reliability and consistency.
- Used numpy, sklearn, and pandas for implementing the models and visualizations.

Infosys Limited — Hyderabad, India
Systems Engineer — Sep 2011-June 2013

- Worked as part of the offshore team for a major aircraft manufacturer on their Data Delivery Program to deliver relevant data according to individual airline plane configuration. Developed a file splitting module to split transaction files for parallel processing of transactions and hence reducing the transaction time by over 50%.
- Performed unit testing of individual components as new features were added and automated them using *JUnit*.
- Worked mostly on *Java, Oracle 9i, IBM Websphere MQ/MB, XML, SQL, Junit, SVN*.

Academic Projects

Predictive Modeling of Drug Sensitivity — Oct 2014-Dec 2014

- Developed elastic nets and lasso regression models to predict drug sensitivity for different cancer cell lines to customize treatment for individual patients. Models and visualizations were developed in R.
- Used Principal Component Analysis for dimensionality reduction as each individual patient data contained 1639 Mutations, 23316 Copy Number and 18988 Gene Expressions.

Google Entity Recognition and Disambiguation challenge — Feb 2014-May 2014

- Participated in the ERD challenge and developed an entity disambiguation model using Clueweb annotated corpus. we developed named entity linking model followed by mention detection and, disambiguation model and further used freebase to extend the lexicon.

Early Alzheimer's Detection — Nov 2013-Dec 2013

- Developed machine learning models using Naive Bayes, SVM and Neural Networks in *Matlab* to classify people based on their performance on a web based test using mouse tracking and clicking data.
- Further optimized the model using convex optimization, feature normalization and optimized bias-variance trade-off.

Technical Skills

- **Languages:** *C, Java, Python, R, Matlab, Pig, HTML/CSS, SQL, XML, Bash scripting.*
- **Libraries & Frameworks:** *pandas, numpy, sklearn, Mallet, OpenMPI, Bootstrap, J2EE, grunt.js*
- **Tools & Platforms**: Git, MongoDB, MySQL, Hadoop.
- **Domain:** Machine Learning, Data Mining, Predictive Analytics, Natural Language Processing.
- **Online Courses:** Udacity A/B Testing and Data Visualization, Stanford Machine Learning.