Puneet S. Ludu

pludu@buffalo.edu **|** 40 Newport pkwy, Apt # 2209, Jersey city, NJ **|** +1-(716) 867-4344

Education

Master of Science in **Computer Science,** State University of New York, Buffalo   **GPA**: *3.482*   *2014*

B. Tech. in **Computer Science,** Jaypee Institute of information Technology, Noida   *2010*

Skills

Skills	Adept	Working Knowledge
Languages	Java, Python, Perl	C/C++, C#, Matlab,
Web Technologies	HTML, Javascript	PHP, Ruby on Rails, jQuery, CGI, JSP
Big-Data Technologies	Apache Pig	Apache Spark, Apache Hadoop, Apache Solr, MongoDB

Independent Projects

Memodiction   An Android app that lets you learn and revise words to build your vocabulary intelligently
ShortCleaner   A JAVA library to clean and normalize tweets for consuming it for various kinds of analytics.
jTextBrew   A JAVA library for fuzzy string matching, based on TextBrew algorithm by Chris Brew
Tweetiments   A simple web based tool to know the sentiment of a particular twitter user (Tweetiments.com)

Experience

NLP Engineer, *April 2015 - Present*, FactSet Research Systems inc., New York City

- Duplicate document identification *(Java, Shell Scripts)*
  - Used *Shingling* and *Vector Space Model* algorithms to detect duplicate documents from a stream of documents.
- Statistical Machine Translation *(Moses, Python)*
  - Building a language translation engine similar to google translate, for Polish, based on principles of statistical machine translation.
  - The cleaning, tokenization, stemming and other language processing is done by using nltk and other NLP libraries in python.
- Earning and Stock Correlation using Social Media Sentiments *(Python, Java, PredictionIO)*
  - Exploratory project to find correlation between social media sentiments with earnings and stocks of a company.

Research Engineer, *July 2011 - July 2013*, TCS Research Labs – Delhi

- **Event Detection in Time Series**   *(Java, Python, Perl, Shell Scripts, Rapidminer)*
  - Worked on a research project involving *multivariate time series with application to classification of events*.
  - Devised an algorithm based on Shape Context for finding frequently occurring patterns and events, with as good results as SAX, DTW etc. with 7% better results in the particular domain of car sensors.
- **Data Harmonization Framework (DHF)**   *(Java, Apache Pig, XML)*
  - Implemented an *ETL framework* that exploits the power of map-reduce and big-databases to fuse incongruous enterprise data from disparate sources in near real time.

Software Development Engineer, *December 2010 - June 2011,* TATA Consultancy Services – Trivandrum

- **Learning Management System** *(PHP, MySql, jQuery)*

Conferences and Publications

- "Inferring latent attributes of an Indian Twitter user using celebrities and class influencers", ACM Hypertext 2015
- "Inferring gender of a Twitter user using celebrities it follows", CORR 2014
- "Architecture for Automated Tagging and Clustering of Song Files According to Mood", IJCSI, 2010

Academic Projects

- **Inferring gender of a Twitter user using celebrities it follows** *(Java, Python, Perl, Google API, Apache Pig)*
This project classifies gender by leveraging observable information such as the tweet behaviour, linguistic content of the user's Twitter feed and the celebrities followed by the user.
- **Quena- A Question and Answering system** *(Java, Apache Solr, Stanford NER, Stanford POS tagger, Apache Pig)*
Indexed 1.6 Million Wikipedia documents, designed a question parser and a ranking algorithm based on popularity.
- **Clustering of songs according to mood** *(Matlab, Perl, C#)*
Using music information retrieval techniques, extracted several features from songs and designed an algorithm, to cluster songs based on moods.
- **English to Hindi Translation and Transliteration** *(C, Gambas)*
Developed tools for Translation using *Hidden Markov Model* and a rule engine engine for Transliteration and word reordering.
- **PSLFS – Virtual File system** *(C)*
A simple file system using K-ary data structure. Won 6th position in ICCC 2008.