

# Shijie Li

(402)-314-9216 | <http://jerry-shijie.github.io> | [sli41@ncsu.edu](mailto:sli41@ncsu.edu)  
1701 Gorman ST 206, Raleigh, NC 27606

## EDUCATION

---

<b>North Carolina State University (NCSU)</b> , GPA: 3.9/4.0 M.S. Computer Science   Supervisor: Dr. Nagiza Samatova, Full professor	Aug. 2015 - Dec. 2017
<b>University of Nebraska-Lincoln (UNL)</b> , GPA: 3.8/4.0 M.S. Physics   Supervisor: Dr. Evgeny Tsymbal, Full professor	Aug. 2011 - May. 2015
<b>Shandong University</b> , GPA: 87/100 B.S. Physics   Supervisor: Dr. Mingwen Zhao, Full Professor	Sep. 2006 - Jun. 2010

## SKILLS AND INTERESTS

---

- Experience in Machine Learning, Artificial Intelligence, Data Mining and Web Development;
- Proficient in *Python*, *Java*, familiar with *R*, *SQL*, *C/C++*, *JavaScript*, experienced with major IDEs;
- Interested in machine learning, data analytics, quantitative modeling and web development positions.

## SELECTED PROJECTS

---

<b>Personalized Recommender System (Java, Python)</b> <ul style="list-style-type: none"><li>• Constructed personalized recommender systems with text analytics and social network mining.</li><li>• Developing feature extraction methods and content-based recommendation algorithms to improve the accuracy.</li></ul>	Sep.2016 - Present
<b>Graph Embedding for Recommender Systems (Python, Linux Shell)</b> <ul style="list-style-type: none"><li>• Created heterogeneous information network with nodes consisting of users, item-ratings, items, and related entities.</li><li>• Implemented DeepWalk algorithm in Python with networkx module to generate random walks over this graph.</li><li>• Based on random walks, embedded the graph in a low dimensional vector space using deep learning model word2vec.</li></ul>	Apr.2016 - May 2016
<b>Matching Algorithm Design for Adwords Problem (Python, MATLAB)</b> <ul style="list-style-type: none"><li>• Simulated Google AdWords online advertising marketing mechanism using random query sequence.</li><li>• Implemented Greedy, Balance and MSVV algorithms to maximize total revenue by advertisers on bid prices and budgets.</li><li>• Obtained best competitive ratio up to 99% in MSVV algorithm.</li></ul>	Mar.2016 - Apr.2016
<b>Market Segmentation using Attributed Graph Community Detection (Python, R)</b> <ul style="list-style-type: none"><li>• Divided social network into segments of common interests based on both vertex attribute similarity and edge intensity.</li><li>• Implemented using Python with the igraph module to facilitate the graph expression and manipulation.</li></ul>	Feb.2016 - Mar.2016
<b>Feature Extraction for Real-time Sentiment Analysis of Twitter Stream (Python)</b> <ul style="list-style-type: none"><li>• Set up data streaming pipeline using Kafka distributed streaming platform, Twitter API and Spark Streaming.</li><li>• Implemented Bag-Of-Words(BOW) sentiment analysis method using Spark RDDs and MapReduce API in Python.</li><li>• Implemented text feature extraction models using Python library module gensim and sklearn.</li><li>• Compared results by BOW and word2Vec using Naive Bayes and Logistic Regression with best accuracy above 80%.</li></ul>	Jan.2016 - Feb.2016
<b>Music Recommender System using Apache Spark (Python, HTML)</b> <ul style="list-style-type: none"><li>• Cleaned and indexed raw data from Last.fm using Resilient Distributed Dataset(RDD) and MapReduce model.</li><li>• Implemented the collaborative filtering recommender model using Spark MLlib in Python.</li><li>• Obtained accuracy of top-K recommendation overlapping rate above 90%.</li></ul>	Dec.2015 - Jan.2016
<b>Library Administration and Management System (Java, SQL)</b> <ul style="list-style-type: none"><li>• Designed E-R model and relational schema for library administration system and implemented in Oracle database.</li><li>• Organized library database using SQL queries embedded in Java code through JDBC APIs.</li><li>• Implemented menu-based UI, time reminder and late penalty messaging components in Java.</li></ul>	Aug.2015 - Dec.2015

## CERTIFICATE

---

<b>Microsoft Professional Program in Data Science</b>	Dec. 2016 - Feb. 2017
---	-----------------------

## TEACHING ASSISTANTSHIPS

---

CSC505: Design and Analysis of Algorithms, CSC230: C and Software Tools. (Fall 2015 - Spring 2017@NCSSU)