**Xi Cui**

Tel.: +1-4695438621; Email: [xxc161130@utdallas.edu](mailto:xxc161130@utdallas.edu)

Link: <https://www.linkedin.com/in/xi-cui-56b886129/>

Github:<https://github.com/cuixi521>

# EDUCATION BACKGROUND

## The University of Texas at Dallas, Richardson, TX Sep 2016- Dec 2019

MS. Major in Statistics (Data science) **GPA**: 3.90/4.0

MS. Computer Science (Data science) **GPA**: 3.81/4.0

## Beijing Jiaotong University, Beijing, China Sep 2010 - Jun 2014

B.E. Information and Computing Science (Applied Mathematics) School of Science

Coursework: Mathematical Analysis, Advanced Algebra, Operations Research, Stochastic process, Probability Theory and Mathematical Statistics, Numerical Analysis, Mathematical Modeling, Graph Theory

# SKILLS

**Language**: R, Python3, Java, C++, SQL, Scala, SAS

**Knowledge**:

Big Data Tool: Hadoop 2.6, APACHE SPARK 2.3.2, AWS

Statistical Methods and Modeling: Analysis of Variance, Model Selection, Monte Carlo method, Advanced Linear Model

Machine Learning and Statistical Learning: Decision Tree, Random Forest, Logistic Regression, Neural Networks, K-means clusters, Adaboosting, Cross-Validation.

Database: ORACLE Sql developer

# EXPERIENCE

## Data Science Internship, BNSF Railway May 2019 – Aug 2019

Data Science. Fort Worth, Texas

* Pull train and car event log data from Customer account database maintenance. Depending on attributes of database, selected target customer sets. Analyzing

client asset distribution, classified customer groups and detected customers’ risk level from assessment data.

* Used Wind Finance Database to track different financial products, gave reasonable prediction and decision advises for

optimal investment portfolio. Implemented data visualization tool to assist explanation and evaluation.

# PROJECTS

## Twitter US Airline Sentiment (Spark MLib on Spark 2.3.2) Nov 2018- Dec 2018

* Created machine learning workflow to classify customer sentiment of US airline company.
* Built and optimized feature engineering pipeline with data acquisition, data pre-processing (spark data frame), term

hashing and model creation (MLib classification).

* Implemented parameter grid builder for tuning parameter and cross validator for finding the best model.

## Search Engine for Movie Plot Summaries (Scala on Spark 2.3.2) Oct 2018- Nov 2018

* Find the top5 most related movie names by giving specific key word.
* Depending on created the key word sets of movies, calculated tf-idf using MapReduce method. Return the top 5 movie name

with highest weight from sorted results.

* Compiled the Scala project into a jar file and applied on AWS cluster.

## Coronary Heart Disease Risk Analyses (R) Jan 2018- Feb2018

* Analysed the dataset “SAheart” to predict whether the patient will get coronary heart disease (“chd” as the response).
* Equipped and compared Logistic Regression, LDA and QDA Model and 10-fold cross validation to get optimal parameter.
* Used the evaluation method of ROC and AUC, result showed that Logistic Regression performed best.

## Walmart Trip Type Classification (Python, Kaggle Competition) Oct 2017- Dec 2017

* Depending on the purchase history of Walmart customers, designed a model to define the type for customers’ trip.
* Adjusted the dataset attributes by preprocessing and feature engineering using Pandas DataFrame to get more reasonable training and testing set.
* Equipped decision tree and neural networks built model and cross validation to improve the classifying accuracy and neural

networks get the lowest log-loss 1.1949.

# WORKING EXPERIENCE

## Full time, China Merchants Bank July 2014 – May 2015

Account Manager. Beijing, China

* Customer account database maintenance. Depending on attributes of database, selected target customer sets. Analyzing

client asset distribution, classified customer groups and detected customers’ risk level from assessment data.

* Used Wind Finance Database to track different financial products, gave reasonable prediction and decision advises for

optimal investment portfolio. Implemented data visualization tool to assist explanation and evaluation.