

WENJIA ZHAI

(420) 890-4973 | zhai.we@husky.neu.edu

OBJECTIVE

Machine Learning Engineer, Data Scientist or Data analysts

EDUCATION

Northeastern University

Boston, MA

M.S. in Bioinformatics, GPA: 4.0

May 2020

Coursework: Statistics, Data Mining, Machine Learning, Data Visualization, Neural Network,
Natural Language Processing

Massachusetts Institute of Technology

Boston, MA

Certification in Data Science and Big Data Analytics

Jan 2019 – Mar 2019

Coursework: Data preparation and transformation, Regression, Classification and Clustering, Neural Network,
Recommendation Systems

TECHINICAL SKILLS

Data Science: Data Mining, Feature Engineering, Regression Models, Classification Models, Clustering Models,
Deep Neural Network, Natural Language Processing, Time Series Analysis, Data Visualization

Programming Language: R, Python, Perl

Programming Frameworks & Tools: TensorFlow 2.0, Keras, SciKit-Learn, Regular Expression,
SQL (MySQL, SQLite), MongoDB, Shell, Git, Google Cloud Platform

Website Development: HTML, CSS, JavaScript, Flask

Operating System: Windows, Linux/Unix (MacOS, Ubuntu)

Report Writing: Microsoft Office (Word, Excel, PowerPoint), LaTeX, Markdown, Jupyter Notebook

ACADEMIC AND WORK EXPERIENCE

Internship: Python Artificial Intelligence Engineer (Intern)

Beijing, China

Huike Group

Sep 2019 – Present

Rationality: The company wants to design a program to automatically segment the online course videos based on the structure of specific videos

- Segmented video based on silence duration based on multiple algorithms
- Built voice recognition system using CTC, CNN and Attention algorithms, transformed audios into texts from videos using TensorFlow 2.0
- Implemented seq2seq algorithm to perform text corrections
- Conducted TextRank model to extract key words in each paragraph

Achievement: The best model achieved 89.2% segmentation accuracy compared to manual segmentation on 20 videos

Project: Extraction and Analysis of Opinions in News in Python

Boston, MA

Team Leader

July 2019

- Collected ~100 news articles from mainstream media websites (CNN, Fox, New York Times, Washington Post)
- Extracted entities of opinions by named-entity recognition using coreNLP and their opinions
- Clustered articles with similar opinions
- Constructed a website using Flask and used D3 to visualize the clusters

Project: Pet Adoption Speed Prediction in R

Boston, MA

Team Leader

Mar 2019 – May 2019

- Performed various types of feature engineering to pets' age, breed, color and other categorical features for data preparation
- Applied linear regression, K-nearest neighbors, decision tree, support vector machine, random forest algorithms
- Evaluated model performance by AUC value

OTHER EXPERIENCE

Co-author

July 2019

- Author of chapter 2: *Smart Search Strategy* in book: *Artificial Intelligence and Natural Language Processing*
- Publisher: China Machine Press (In Progress)