

# Zidong Xu

Washington, DC 20007 | (202) 644-2783 | [zidong2020@gmail.com](mailto:zidong2020@gmail.com) | [www.linkedin.com/in/zidong-xu](http://www.linkedin.com/in/zidong-xu) | <https://zidong2020.github.io/show-case/index.html> (more project display)

---

## EDUCATION

### GEORGETOWN UNIVERSITY

Washington, DC

#### Master of Data Science and Analytics (STEM field)

Expected Graduation: May. 2023

- GPA: 4.0 / 4.0

### CENTRAL CHINA NORMAL UNIVERSITY

Wuhan, China

#### Bachelor of Information management

Sept. 2016 – Jun. 2020

- GPA: 3.61/ 4.0 (top 5%)

---

## EXPERIENCE

### GOUSEN SECURITIES CO., LTD

Beijing, China

#### Department of Fixed Income, Analyst Intern

Mar. 2021 – Aug. 2021

- Developed an understanding of economic and political forces affecting the issuance of bonds and the bond market through internal and external resources.
- Prepared three due diligence documents for bond issues (projects between 1 billion and 2 billion).
- Developed data and analytical processing skills, proficient knowledge of available computer programs & improved data processing systems to enhance departmental efficiency.
- Developed the company's customer service system, assisting clients in preparing and publishing annual and audit reports.

### Trip.com Group Limited

Shenzhen, China

#### Air Ticket R&D Department, Assistant data analyst

July. 2019 – Aug 2019

- Gathered data from various sources (Social media, News, Official websites) and analyzed information necessary to prepare client presentations, offering statements, and marketing materials (SQL, Excel, PPT).
- Wrote two analysis reports using (English and Chinese) and showed them to the WeChat public account.

---

## PROJECTS

### Big Data and Cloud Computing: Analyze Reddit data with Azure

Sept. 2022 – Nov 2022

- Performed exploratory data analysis (EDA) on two subreddits datasets, submissions (1770430 rows) and comments(38152167 rows), and used WordCloud/matplotlib/plotly to display information about variables.
- Built Pipelines, performed text analysis, and created dummy variables with Spark NLP and NLTK.
- Built Machine Learning Pipelines (ML Spark), such as Random Forests/Logistic Regression, to select essential features and evaluate models' performance (Accuracy, Confusion Matrix, and AUC).

### Natural Language Processing (NLP): DBpedia14 classification with BERT

Sept. 2022 – Nov 2022

- Did exploratory data analysis (EDA) and Data preprocessing (solving uneven distribution labels of the dataset).
- Created batched inputs using Huggingface's DistilBERT tokenizer.
- Fine-tuned Huggingface's DistilBERT model, Trained the DistilBert model on the training set. Computed and printed out the trained model's loss, accuracy, and f1-score on the train and test sets (Use Google Colab+, GPU).
- Visualized embeddings using sklearn TSNE.

### Neural Nets and Deep Learning: Object detection and time series analysis

Sept. 2022 – Nov 2022

- Downloaded the Video from Youtube and used OpenCV to read Video and convert it into an ordered set of images.
- Used YOLOv5 to detect vehicles from the current images (frames).
- Created new variables according to the information from the model result and performed a time series analysis.

---

## DISTINCTIONS

**Languages:** English, Mandarin

**Skills:** Python, R, MySQL, [Tableau](#), VS Code, AWS, Azure, Hadoop, Spark, NLP, Time series analysis, HTML

**Packages:** PyTorch, TensorFlow, Scikit-Learn, Matplotlib, plotly, ggplot2, OpenCV, NLTK, Gensim, Streamlit

**Certification:** The securities qualification certificate (China)