Yanming Liu

Email : yanmingliu21@gmail.com
Mobile : +1-617-959-5728

Education

- **Northeastern University** — Boston, USA
  Master of Science in Data Analytics Engineering; GPA: 3.92 — *Expected May 2023*
- **Shandong University** — Jinan, China
  Master of Science in Mechanical Engineering — *June 2020*
- **Qilu University of Technology** — Jinan, China
  Bachelor of Science in Mechanical Engineering — *June 2017*

Related Courses: Algorithms,Database Management for Analytics, Computation and Data Visualization, Data Mining in Engineering, Machine Learning/Current, Neuron Networks and Deep Learning/Plan

Publications & Conferences

- **Zhang S, Liu Y et al.**: Dynamic relationship between noseleaf and pinnae in echolocating hipposideros bats, 2019
- **Liu Y, et al.**: Pattern Matching between Sequence Trains and Structural Deformations in Roundleaf Bats, 2021
- **Liu Y, et al.**: Overview of PVD Coating and Its Life, 2018
- **"Pulse-train time structure for dynamic biomimetic sonar"**: 177th Meeting, Acoustical Society of America. Louisville, Kentucky, 13-17 May 2019
- **"Learning the relationship between biosonar pulse trains and peripheral dynamics in hipposiderid bats"**: 178th Meeting, Acoustical Society of America. San Diego, California, 2–6 December 2019

Technical Skills

- **Languages**: Python, SQL, Java, R, Matlab, LaTeX
- **Database**: Conceptual EER Modelling, Normalization, PlantUML, Cypher, MongoDB
- **Data Process and Analysis**: Numpy, Pandas, Sklearn
- **Data Visualization**: Matplotlib, Tableau, Office Excel, Flourish, Data Wrapper, Google Site

Work Experience

- **Northeastern University** — Boston, MA
  Teaching Assistant — *January 2022 - Present*
  - **Tutoring**: Enriched student learning experience via holding essential SQL training sessions
  - **Advising**: Supplied instant assistance in lecture and fixed database-related issues, 20 hours per week
  - **Grading**: Graded projects, assignments, quizzes, and final exam for nearly 55 students

- **Zhangmen Education Inc.** — Qingdao, Shandong
  Tutor — *July 2020 - December 2020*
  - **Tutoring**: Taught Physics, Mathematics, and English for 20+ junior high school students online one to one
  - **Advising**: Planed teaching to students' needs, and answered students' questions
  - **Optimizing**: Researched in conferences and training activities to improve class quality

- **Shandong-Virginia Tech Int. Lab.** — Jinan, Shandong
  Administrative Assistant — *June 2019 - July 2020*
  - **Reception**: Facilitated the visiting professors & students from the US and presented academic reports
  - **Operation**: Managed lab affairs with University Administration, and Finance Department regarding operations
  - **Collaboration**: Cooperated with researchers sponsored by IRES summer program at Virginia Tech on field experiments

Research Projects

- **Labor and Industrial Skills Gap Analysis in Healthcare** — Northeastern University
  Principal Investigator, Thesis — *May 2022 - Present*
  - **Research Objective**: Identifying labor and industrial skills gap in Healthcare 4.0 sector from 2017 to 2022 in the US as an reference for job orientation and skill enhancement of individuals who target at starting or switching occupations
  - **Existing Problems**: The related research have three main weaknesses: 1) Limited data sets are not persuasive to convey inspiring insights for labor and skills gap in Healthcare; 2) Rarely current analysis keep track of the most recent job and skill trend; 3) Hardly correlated supply and demand investigation for data-related occupations and skills in Healthcare 4.0;
  - **Research Framework**: Present demand, supply, gap of labor for Healthcare occupations on region and nation levels in the US; Further address the labor gap trend for years; Regional distribution for lab gap; Compensation for occupations; Identify the essential skills, including domain knowledge, technical skills(traditional, data-related) and soft skills, for Healthcare 4.0; Unfold skills gap among Healthcare careers and supply/demand for each occupation
  - **Potential Data**: Standard occupational classification codes(SOC); Metropolitan statistical areas(MSAs); Traditional labor market information(LMI); Classification of Instructional Programs(CIP); Job posting analytics(JPA) & Profile data(PD);
  - **Project Management**: Establish research plans, segment research tasks for partners, track research progress, and summarize learning outcomes
  - **Research Expectation**: Publish at least one journal paper, compose master's thesis on the insights gleaned from the labor data, job posting and personal profile data, and offer advices for future smart curriculum recommendations in terms of lagging skills or domain knowledge

- **Dynamics of Bat Sonar Sensing and Physical Mechanism of Coupling Effects** — Shandong University
  Participant, Thesis — *September 2017 - June 2020*
  - **Project Fund**: National Natural Science Foundation of China(11374192), under this
  - **My Research Objective**: Investigated whether there is a relationship between the two sources of variability, the noseleaf and pinna motions with the spiking timestamps of biosonar pulses using a novel SSIMS framework, which was originally developed for mapping the kinematic information and its neuron state in primates
  - **Literature Review**: Set up research direction under reviewing hundreds of investigated publications on bats
  - **Data Collection**: Adjusted high-speed cameras and ultrasound microphone arrays to capture the motions located in noseleaf and pinna of Pratt's roundleaf bats along with the emitted ultrasonic pulses trains
  - **Analysis Framework**:
    1) Noseleaf/Pinna structure in Hipposideros pratti and their motion metrics were introduced;
    2) Qualitative description and quantitative classification of noseleaf and pinna coupled motions were conducted;
    3) Firing times of 102 ultrasound sequences under two aforementioned synergistic motions were addressed;
    4) SSIMS analysis framework using V-P distance metric and t-SNE dimensionality reduction was applied on firing times for classification being evaluated by pattern recognition accuracy among hyper-parameters;
  - **Research Output**: Wrote master's thesis, journals, abstracts as aforementioned in section Journals & Conferences and presented research results at two international conferences.
    Insights are as follows:
    1) These joint kinematics data on the noseleaf and pinnae have shown both qualitative and quantitative relationships between the noseleaf and pinna motions: large noseleaf deformations (opening or closing) tend to be associated with non-rigid pinna motions. Furthermore, closing deformations of the noseleaves tend to co-occur with closing motions of the pinna. Finally, a canonical correlation analysis of the motion trajectories has revealed a tight correlation between the motions of the landmarks on the noseleaf and both pinnae;
    2) Variability in biosonar pulse sequence match the peripheral dynamics, i.e., different noseleaf/pinna motion patterns correspond to the biosonar pulse sequences with different pulse numbers and emitting times;

- **State Transition in Swarms of Bat** — Shandong University
  Participant, Field Research — *September 2017*
  - **Field Record**: Conducted data collection on cave's location, bats' living environments, and its quantity
  - **Experiment Design**: Designed experimental plan, team delegation, and in charge of equipment set up
  - **Report Output**: Summarized and communicated experiment result daily to project supervisor

Course Projects

- **Prediction of Heart Disease with Emphasis on Factors Impacting it** — Northeastern University
  Participant 1/2, Data Mining — *January 2022 - May 2022*
  - **Objective**: Conducted Data exploratory analysis and optimal machine learning model choice on heart disease data set in Kaggle using Python to improve quality of life and treat heart disease in the earlier stages for individuals
  - **Data Source** & **Explanation**:
    70000 records, 11 attributes and one target column for disease presence or not. Among attributes, 6 features are categorical and the rest are numeric. All features can be categorised into 3 parts, objective, examination and subjective: Age, Height, Weight, Gender | Systolic and Diastolic blood pressure, Cholesterol and Glucose level | Smoke, Drink, Exercise Status
  - **Data Preprocessing**: Outliers were eliminated by domain knowledge on summary statistics and pairplot for numerical variables
  - **Exploratory Data Analysis**: The quantity distribution of numerical variables versus heart disease status(boxplot); The frequency of each category in categorical variables versus heart disease status(bar chart); Correlation analysis and Chi-square method for features
  - **Data Mining Models**: According to the classification metrics, chose the best model among K-Nearest Neighbors(KNN), Logistic Regression, Decision Trees, Random Forest, and Neural Networks
  - **Output**: Report, Presentation, and Ipython file for data processing, analysis, models choice. Insights are as follows:
    Compared to the KNN, Tree-based method and Logistic Regression, the optimal accuracy of 73.47% is obtained when training a Neural Network under the best parameters, where the parameters used were 2 hidden layers with 5 nodes each and activation function being logistic and solver being lbfgs. The second highest is the Decision Tree with an accuracy of 73.2%. Glucose, Cholesterol and Blood pressure have a significant impact on heart disease. Overall, Neural Networks prove to be the best among all which would be used as our standard model for the application that the general population can use.

- **Factors Analysis of Impacting Alcohol Consumption and Grades** — Northeastern University
  Participant 1/3, Analysis & Visualization — *September 2021 - December 2021*
  - **Objective**: Investigated student audience, factors impacting alcohol consumption and grades in Mathematics class of Portuguese Schools using R
  - **Data Source**: Student Performance Data Set
  - **Analysis Framework**:
    1) Exploratory data analysis was conducted on the dataset to understand the dataset, looked for points of interest and determined what portion of the data should be explored next;
    2) Alcohol consumption habits of Portuguese students were compared against demographic and social indicators such as age, parents' jobs, social life in analyzing factors impacting alcohol consumption among math students in Portuguese Schools;
    3) Geographical, social and demographic data points like study time, travel time, health, and internet access were compared to find trends affecting students' grades and the aforementioned analysis on alcohol consumption was used to determine relationship between students' grades and alcohol consumption
  - **Report Output**: Using R Markdown to compile the description and analysis for report

- **Demo: Bank Management Database System Design and Query** — Northeastern University
  Principal Designer 1/2, Database — *September 2021 - December 2021*
  - **Objective**: Understood the framework from the perspective of the business problem, database structure, implementation, and query in SQL and NoSQL
  - **Database Design**: Designed conceptual model and logical model on need of investing, saving, loaning for customers; the financial product purchase between the bank and 3rd party company; Customer support services; Human resources management
  - **Database Implementation** & **Query**:
    Created SQL and NoSQL databases with 20 artificially generated records, and queried questions with DML, MongoDB, Cypher languages under Python. This questions can be company profit, potential customers for investment with higher deposits, top performing employees, etc.
  - **Output**: Wrote project report, presented for 15 minutes, and got the teaching assistant job for Database Management in Analytics