ZHAO Fuxian

+86 13511515376 | 211505013@smail.nju.edu.cn
Nanjing, Jiangsu, China

EDUCATION

Nanjing University(NJU) Sep 2021 - Jun 2025
Bachelor of Science in Bioscience Nanjing,China
- GPA: 4.34/5
- Honer: The People's Scholarship in China, Undergraduate Basic Subjects Special Scholarship
- Python Programming(90), Calculus II(91), Biostatistics(86), Data Mining(89), Database and Information System(86)

City University of Hong Kong(Cityu) Sep 2023 - Dec 2023
Exchange program in Mathematics Hong Kong
- GPA: 3.73/4.3
- Applied Statistics for Science and Engineering (A), Discrete Mathematics(A), Python for Data Science(A), Fundamental of Machine Learning(A-)

WORK EXPERIENCE

NIO Inc. Apr 2024 - Jul 2024
Data Analyst(intern) Shanghai, China
- Utilized sql and tableau to monitor core community-related metrics on a weekly basis, and explained data anomalies through drill-down analysis.
- Through correlation analysis and linear regression, analyzed the relationship between user value and activity of vehicle owners with different identities and registration times to help the operation team achieve personalized operation
- Utilized Correlation coefficient matrix and PCA to extract principal community performance features , and determine the classification of car owners based on kmeans algorithm

TAL Inc Jan 2024 - Apr 2024
Data Analyst(intern) Beijing, China
- Leveraged SQL and proprietary data platforms to create and maintain real-time data dashboards, monitoring KPI metrics such as student course completion rates and subscription renewals
- Autumn and Winter 2023 Core Data Insights: Identified key trends and patterns to drive curriculum improvements.
- Winter Special Courses Correlation Analysis: Investigated relationships between various instructional metrics to optimize course delivery.

PROJECT EXPERIENCE

Data Science Salary Prediction Project Dec 2023 - Dec 2023
City University of Hong kong Hong Kong
- Predicted salary categories usingmachine learning classifiers(Logistic Regression, Random Forest, and SVM).
- Web scraped job description data and transformed text data into continuous and binary variables.
- Conducted data cleaning, exploratory data analysis and feature engineering, improving data quality for model training.
- optimized model performance through cross-validation and hyperparameter tuning to avoid overfiting, achieving a balanced accuracy of 73% with SVM model.

Data Science Summer School Jul 2023 - Aug 2023
Imperial College London, UK
- CNN-based Brain Tumour Segmentation Network:
  - Data augmentation: conducted crop, random horizontal flip and random vertical flip on training data
  - Enhances the traditional U-Net model: introduced batch normalization after convolution , random dropout and an embedding layer and chose binary cross-entropy loss and soft IoU loss as loss function
  - Achieved Best Computer Vision Project award for innovative model improvements and high performance.
- Word Representation in Biomedical Domain:
  - Parsed the data: processed the original json files which store Covid-19 literature, eliminated non-word text.
  - Tokenization: generated tokens by the version of Scispacy specifically for biomedicine area
  - Built Word Representation based on N-gram and Skip-gram respectively
  - Explored results : reduced word vectors by t-SHE and visualize biomedical words

Genome-wide Z-RNA Prediction Feb 2023 - Jul 2024
Nanjing University Nanjing, Jiangsu, China
- Set up a control group with APOBE2 and APOBE3 added respectively and an experimental group with APOBE connected to Zaa1 and Zaa2 respectively
- Editorial site statistics: mainly counted the proportion of edited sites in A to I and C to U in each group and compared editing in alu and non-alu regions
- Validation of purine-pyrimidine motifs: conducted kmer analysis for clusters of genes where site editing occurs, counted kmer 6 to 12 complexities defining front and back drift zero pyrimidine interleaving as a complexity

Bank Customer Behavior Analysis and Churn Jan 2023 - Jan 2023
Nanjing University Nanjing, Jiangsu, China
- Completed the relevant literature review
- Dimension reduction: identified principal variables through neural network model
- Used XGBOOST and Random Forest model to predict respectively, and finally found the key to bank customer churn lies in recent assets, deposits

Skills

- Programming/Tools: Python (Pandas, Matplotlib, Seaborn, Scikit-learn,pytorch), SQL, Tableau
- Data mining: Statistical Analysis, Predictive Modeling, Data Visualization
- Machine Learning: classic regression and classification algorithms , cluster algorithms , PCA , CNN model