**Xinzhou Wang**

Email: xinzhou\_wang@outlook.com· Tel: (+86) 18221546090

**education background**

**Shanghai University – The University of Technology Sydney (SHU-UTS), China** 09/2020–06/2024

B.Mgt. in Info Management and Info Systems | Dual Degree: B.Bus. in Management | Overall GPA: 3.55/4.0

**Coursework (Math & Programming):** Multivariable Calculus | Linear Algebra | Probability and Statistics | Data Structure | Advanced Program Design | Algorithm Design and Analysis | System Analysis and Design | Financial Programming and Computing | Database | Data Analytics | Data Warehouse | IT Project Management | Data Science | Game Theory

**Coursework (Business):** Economics | Financial Accounting | Management and Organization | Marketing | Operations Research | Global Operations and Supply Chain Management | Human Resources Management | Managing People and Organizations

**awards and honors**

*The 2nd Prize* | 2021&2023 Shanghai University Academic Excellence Scholarship

*The 2nd Prize* *(National Final)* | 2021 IEERA China Business English Debating Competition

*Outstanding Student Leader* | SHU-UTS Student Union

*Best Leadership Award* | Shanghai University International Students Association

**skills**

**Programming Languages:** C/C++ | Python | MATLAB | R | Stata | SPSS | SQL | JavaScript | VBA

**Libraries & Frameworks:** NumPy | SciPy | Pandas | Sklearn | Seaborn | Pygame | Selenium | PyTorch | TensorFlow | Django

**Software:** Linux Ubuntu & CentOS | MS Office | PyCharm | Visual Studio | Kettle Spoon | MySQL Workbench | PowerBI

**work experiences**

***Data Analyst Intern |* Executive Office, X.D. Inc. China**  12/2023–06/2024

* Managing, processing, and analyzing millions of user’s data, leveraging business intelligence (BI) tools such as Tableau and Power BI for visualizing data; forming and presenting comprehensive and in-depth user research reports to guide the follow-up game operations and development, and ensuring that strategic decisions are backed by empirical evidence derived from complex data analyses involving multivariate testing and cohort analyses.
* Assisting data engineers by leveraging robust database and data warehouse skills to 1) create a structured data flow, including data integration tools and visualized data pinboard, for a non-trivial newly-published mobile game, 2) design and create a data warehouse diagram, data dictionary, data snapshot (synchronization logic) by using kettle Spoon based on JavaScript, etc. 3) enable analysts to acquire structured or visualized data conveniently, minimizing the dependency on SQL queries and enhancing the data retrieval process.
* Designing and developing a reusable NLP system that stratifies and rates players’ reviews. This workflow that enables sentiment analysis of players’ attitudes towards our game product done by machine learning technologies such as Naive Bayes, SVM and statistics methods such as TF-IDF significantly facilitates data analysts in the forming of comprehensive business reports.

***Business Data Analyst Intern* | Tencent Technology (Beijing) Co., Ltd. China**  06/2023–07/2023

* Built data-driven workflows to promote the engagement and retention of customers:
  + Enabled 1) customer segmentation via k-means clustering, 2) (NLP) customer review analysis and sentiment analysis via cross-validated random forest, gradient boosting, logistic regression with texts TF-IDF encoded, and Naïve Bayes models, 3) customer preference prediction via random forest, boosting, and GLM, and 4) personalized service recommendation via collaborative and content-based filtering.
  + Helped secure, harmonize, and organize large-scale spatial and non-spatial customer data with the aid of Nginx, Hadoop, Kafka, Flume, Flink, and Elasticsearch libraries, frameworks, and platforms; used domain-specific language (DSL) to drive Elasticsearch query and analysis; performed data visualization in Kibana.

**software development projects**

**Development of an Integrated AI Voice Assistant App for a Tunnel Construction Site** 07/2023–09/2023

* Designed, implemented, tested and deployed an AI Voice Assistant for application in a tunnel construction site:
  + Reviewed and characterized a set of open-source modules for automatic sound recognition (ASR) and text-to-speech (TTS) conversion, with a focus on underlying machine learning models (e.g., MASR framework and gradient boosting decision tree); analyzed the usability, robustness, performance, extensibility, and maintainability of the modules from software engineering perspectives; identified the optimal modules for app integration.
  + Built and integrated five functional modules, encompassing audio recording, voice recognition, chat response, TTS execution, and sound playback; leveraged OpenAI and ChatGPT to drive model development; completed a rigorous test using an array of metrics, such as word error rate, missed error rate, word information loss, etc; deployed the product to Linux Ubuntu and Windows environments.
  + Delivered a comprehensive report documenting the design principles, testing methodology, and system demo.

**Development of a Personal Blog Website (**[**Link**](http://124.222.120.214/index/)**)** 12/2022–Present

* Designed, implemented and deployed a personal blog website based on Django web framework and Nginx web server:
  + Collectively used Python, Django, Nginx, Hadoop, Kafka, Flume, Flink, and Elasticsearch libraries, frameworks, and platforms to drive development, testing and maintenance, covering frontend and backend.
  + Followed a set of software engineering principles and design patterns to promote efficiency, modularity, extensibility, maintainability and robustness.

**Development of a Data Warehouse for PC Game Distribution and Profitability Analysis** 11/2022–12/2022

* Designed, prototyped and validated a database system based on MySQL to support the distribution of PC games, with a focus on data warehouse construction for data logging and profit analysis, allowing data-driven strategy optimization
  + Built ER diagram accordingly to define relation between different elements; analyzed functional dependency, and performed Boyce-Codd normal form (BCNF) decomposition and third normal form (3NF) decomposition; gained familiarity with the design, implementation and testing of a relational database system with the aid of MySQL Workbench GUI for efficiency, security, and robustness
  + Established a data pipeline based on extract, transform, and load (ETL) process, covering raw data collection, cleansing, aggregation and saving; tested the database against 20K+ transaction history data points, covering a broad range of e-business queries; enabled data visualization using PowerBI
  + Integrated a data-driven customer profiling and clustering workflow for identifying upselling and cross-selling opportunities (i.e., inviting customers to purchase high-end and complementary PC game products, respective)
    - Trained logistic regression, GLM, and decision tree models for customer profiling and clustering
    - Used ROC/AOC metrics and F, recall, precision, accuracy, and confusion scores to evaluate performance.

**Prediction of Taxi Trajectories Based on Geolife GPS Dataset** 10/2022–11/2022

* Designed, implemented, and tested a data-driven workflow for visualizing and predicting the taxi trajectories in a metropolitan setting, allowing dispatch guidance for drivers and effective traffic management:
  + Completed literature review and market survey for problem identification
  + Enabled data extraction from Geolife GPS V1.3 from Microsoft Research Asia, followed by 1) systematic data cleansing to reconcile data format, reduce data noises and remove anomalies, 2) data segmentation and sematic augmentation, 3) dimensionality expansion, and 4) feature encoding, normalization and selection; prepared training and testing datasets to support model construction
  + Leveraged various geospatial data visualization techniques to facilitate pattern extraction and result presentation
  + Trained, validated and tested a set of supervised learning models to drive prediction, covering KNN, multiple linear regression, SVM, random forest, and LSTM models; systematically evaluated and compared model performances using fit-for-purpose metrics

**Project Management for Enterprise Informatization – Issue Identification and Problem-Solving**

***IT Project Management* Course Project, SHU-UTS** 09/2022–11/2022

* Completed a case study on an enterprise informatization project, comprehensively covering problem identification and strategy design for scope, progress, cost and team management:
  + Performed Monte Carlo simulation for predicting project progress, covering various scenarios; leveraged critical path method (CPM) to drive progress management
  + Enabled cost estimation using various techniques, e.g., Walston-Felix model, EVM, Delphi methods.

**A Data-Driven Study on the Factors Affecting the Flight Delay** 05/2022–06/2022

* Completed a data-driven study to identify the key factors dictating the delay of American Airline flights:
  + Mined real-world data from various online sources using Selenium and WebDriver; visualized the initial dataset to assess its distribution, balance and completeness; leveraged the relevant domain knowledge to identify new features
  + Conducted data preprocessing in Pandas to identify independent variables and reduce feature dimension, covering multicollinearity test, heteroscedasticity test, and stepwise regression.
  + Performed exploratory data analysis on the raw dataset to extract patterns, identify anomalies, test hypothesis, and validate assumptions; identified explanatory variables and dependent variables
  + Employed logistic regression, binomial and multiple classification tree models to drive study

***Financial Programming and Computing* Course Project, SHU-UTS** 04/2022–05/2022

* Built an event-driven stock trading platform to guide stock selection, portfolio optimization and transaction decision-making, covering Python and MATLAB modules for
  + Updating technical indicators, including overlays (moving average, Bollinger Bands) and oscillators (stochastic oscillator, MACD, RSI).
  + Calculating momentum alpha factors for stock selection, covering OPSD, alpha13, alpha55, ADX, annual firm set growth rate, CO, turnover return, Bias turnover, intraday minus overnight expected returns, etc.
  + Performing automated stock trading based on calculated technical indicators and various strategies, including pairs trading, grid trading, cross market trading and trend trading, followed by a backtest to evaluate the performance.

**Design and Implementation of a Shooting Game** 09/2021–11/2021

* Designed, and implemented in Pygame a shooting game, covering a set of weapons, level layouts and layered visuals:
  + Leveraged Unified Modeling Language to guide software design; optimized various components in the game design to enhance the player experience, covering mechanics, feedback, pacing and interface.
  + Built a fit-for-purpose 2D physics engine to model object collision, allowing collision detection, contact manifold generation, physical response, contact caching, and constraints modeling.
  + Enabled team collaboration and version control using GitHub.

**extracurricular experiences**

***President* | International Dept., Shanghai University** 07/2020–09/2022

* Led a team of student volunteers to provide support to the international student affairs, serving a group of 200+ international students and faculty members.
* Coordinated a set of events to promote Shanghai University in overseas communities; helped establish and maintain public accounts on various social media platforms, such as YouTube, Instagram and WeChat.