Yuexin(Cindy) Chen

(336) 473-8093 | cindychen666777@gmail.com | Evanston, IL

EDUCATION

Northwestern University **Evanston, IL**
M.S. in Analytics *Expected Dec 2023*

- **Honors**: Applied Data Science Fellowship, GPA: 3.93/4.00
- **Relevant Coursework**: Analytics for Big Data (Hadoop/Spark/Hive), Cloud Engineering (AWS, A/B Test, Docker), Deep Learning, Predictive Analytics

Wake Forest University **Winston-Salem, NC**
B.S. in Mathematical Statistics, B.S. in Computer Science *May 2022*

- **Honors:** Summa Cum Laude (GPA: 3.95/4.00), 2022 Outstanding Student in Statistics Award
- **Relevant Coursework**: Machine Learning | Statistical Inference | Data Structures and Algorithms | Software Engineering
- **Leadership:** Statistics Department Teaching Assistant, Intercultural Ambassador, Computer Science Mentor

SKILLS

- **Programming:** SQL | Python (NumPy, Pandas, Scikit-Learn, SciPy, Seaborn, TensorFlow, nltk) | R | Java
- **Software**: Jupyter Notebook | MySQL Workbench | Tableau | Visual Studio Code | IntelliJ | Git | Microsoft Office | RStudio

PROFESSIONAL EXPERIENCE

Data Science Graduate Consultant *Sep 2022-Jun 2023*
Honda R&D Evanston, IL

- Transitioned the company's project cost prediction approach from manual calculations to machine learning models such as **Linear Regression, KNN,** and **Decision Tree**, resulting in an average 96% reduction in MAE.
- Developed 6 **Tableau** dashboards with integrated data modeling and prediction system and trained PM on using the tool,
- Engaged with multiple departments to streamline monthly budget allocation and designed an automated budget calculation Excel interface with VBA.

Data Analyst Intern *June 2021-August 2021*
WeBank Shenzhen, China

- Constructed **ETL** pipelines with data cleansing using **SQL** and **Python** with 2M+ loan transaction data to enable data segmentation and improve infrastructure for analytical needs
- Created 7 **Tableau** dashboards with interactive filtering tools on quarter, region, industrial sector, and marketing channel to facilitate historical loan transaction query and visualization; tutored fellow analysts on how to use them for query
- Automated marketing channel conversion reports and loan funnel visualization using **Tableau**; saved fellow analysts 2 hours weekly for generating plots and manually sending emails
- Performed predictive modeling (**support vector machine, random forests**) in Python and **survival analysis** to predict customers' risk scores and assign interest rates for new customers based on their credit reports and personal information

PROJECT EXPERIENCE

End-to-End ML Portal for Heart Disease Prediction **Evanston, IL**
Cloud Engineering Final Project *April 2022*

- Designed an end-to-end machine learning pipeline on **AWS** for heart disease prediction, managing data ingestion, preprocessing, and model training using AWS services (**S3, Lambda, ECR, API Gateway, CloudFront**).
- Applied grid search for optimal hyperparameter tuning while fitting a **Random Forest** classification model, ensuring effective model performance and prediction accuracy.
- Established a model version control system in an AWS S3 bucket for efficient retrieval and real-time prediction, enabling a robust serverless inference architecture with AWS Lambda and API Gateway.
- Developed an interactive, user-friendly web interface with **JavaScript, HTML,** and **CSS**, providing personalized heart disease predictions and health recommendations.

Examining Patients' VISI Posture Recordings at Wake Forest Baptist **Winston-Salem, NC**
Student Researcher for Honor Project, Wake Forest University *Aug 2021-May 2022*

- Cleaned and performed feature engineering through **R** cloud computing on 3,000+ patients' postures datasets and used smoothing and error correction techniques to improve data quality for analytical and research needs
- Conducted exploratory analysis to visualize patients' daily posture behaviors using **R-ggplot2** and **Python-Matplotlib**
- Computed interpretable patients' summary statistics and built intermediate tables for individual patients for deep analysis with **R**
- Performed a clustering analysis using **K-Means++** method and identified 5 posture habit personas; analyzed and modeled the posture impact on the length of hospital stay and developed a plan for healthy patient posture habits for clinicians