CHUNLIN HE, PhD

Data Scientist, Microsoft Certified

chhe09@gmail.com • (765) 490-7331 • Lafayette, IN • <u>LinkedIn</u> • <u>Github</u> • <u>Website</u>

SUMMARY

Results-driven Data Scientist with a Ph.D. in Plant Science, specializing in Biostatistics, and holding an MSc in Biostatistics, along with Data Science and Professional Certificates from Microsoft. Proficient in Python, R, SQL, and SAS, demonstrating expertise in Machine Learning and Deep Learning. Skilled in optimizing predictive models, conducting thorough assessments, and refining models to enhance business performance. Experienced in data wrangling, statistical analysis, and adept use of various data science tools. Possessing strong experience in managing end-to-end data projects, implementing relational database management systems, and conducting data analysis on medical science datasets. Successfully completed multiple statistical and quantitative research projects with peer-reviewed publications. A dynamic problem-solver, poised to excel in data science projects.

EDUCATION

- Data Science Certificate: Microsoft Professional Program (MPP) for Data Science, Microsoft, USA
- Professional Certificate: Data Science Fundamentals, Microsoft, USA.
- Ph.D.: Plant Science with a focus on Biostatistics, University of Saskatchewan, Canada.
- M.Sc.: Biostatistics and Quantitative Genetics, Nanjing Agricultural University, Nanjing, China.
- B.Sc.: Crop Science, Hunan Agricultural University, Changsha, China.

SKILLS

- Programming Languages: Python, R, SQL, KQL, SAS, JavaScript.
- Machine Learning (ML) and Deep Learning (DL): Linear Regression, Logistic Regression, Random Forest, Decision Tree, SVM, Gradient Boosting (AdaBoost, XGBoost, CatBoost), KNN, Neural Networks (computer vision, image analysis).
- Data wrangling and statistical analysis: EDA, Hypothesis Testing, Correlation and Regression Analysis, Chi-square Test, ANOVA, t-test, Statistical Modeling, Principal Component Analysis, Nonparametric Statistics.
- Tools and Technologies: Github, AzureML, Power BI, Spark, Tableau, HDInsight, SQL Server Management Studio, RStudio, MySQL, Kusto.Explorer, Jupyter Notebook, VS Code, PyCharm.

PROFESSIONAL EXPERIENCE

Modeller/Data Scientist, Tata Consultancy Services Ltd., Lafayette, IN　　　　**2020 - 2023**

- Conducted comprehensive assessments of predictive models, identifying gaps for improved model risk management and business performance.

- Specialized in developing and refining models related to business operations, with a focus on predicting customer satisfaction and membership retention.
- Evaluated model accuracy, weaknesses, and limitations through rigorous testing and benchmarking, resulting in significant improvement in predictive performance.
- Implemented effective model maintenance, monitoring, and documentation practices, contributing to continuous business performance improvements.

Microsoft/MPP Data Science Track & Freelance Data Scientist, Lafayette, IN   2019 - 2020
- **Health Industry Collaboration (NDA):** Conducted data mining and database management for a private health company, ensuring data integrity and accuracy. Implemented data cleansing, query operations, and merging processes for large data sets.
- **Clinic Data Analysis (NDA):** Analyzed extensive datasets from a private research organization using R, offering valuable insights derived from statistical analysis. Data sets aggregated from tens of thousands of participants supported decision-making and enhanced organizational outcomes.
- **Relational Database Management (RDBMS):** Implemented a successful project for a private company, establishing a relational database management system (RDBMS) using MS SQL Server. Executed data manipulation, storage, and retrieval operations.
- **Microsoft/MPP Machine Learning Projects:** Demonstrated expertise through the successful completion of diverse machine learning projects including classification and regression. Addressed real-world data science challenges commonly faced by corporations and research organizations.

Alfalfa Toolbox/Data Curator, Noble Research Institute, Ardmore, OK   2015 – 2018
- Data Processing: Organized and manipulated diverse raw datasets previously collected from multi-environmental field trials. Performed exploratory and statistical data analyses to ensure quality control and assurance from a statistical perspective, i.e. statistical inference for entries/treatments across experiments, blocks, locations and years.
- Data Curation and Management: Led data-driven decision in curating and utilizing the qualified field data and corresponding genomic data for integration into the Alfalfa Toolbox, including end-to-end data pipeline for molecular breeding acceleration.
- Toolbox Enhancement: Collaborated with team members in integrating various datasets into Alfalfa Toolbox, facilitating navigation and retrieval of data for seamless access and utilization across alfalfa breeding communities.

Breeding Services Manager/Senior Scientist, GCP, c/o CIMMYT, Texcoco, Mexico   2010 - 2014
- Spearheaded and coordinated over 100 data projects, overseeing the generation of genomic and genotypic data for 100+ academic and research institutions spanning 50+ developing countries.

- Applied advanced data analysis techniques, such as QTL mapping, to extract actionable insights that facilitated the integration of genotypes and phenotypes for international breeding programs.
- Mentored and guided a team of research scientists, providing expertise in utilizing cutting-edge data analytics tools and creating visualizations for the effective analysis of genotypic and phenotypic data.

REFERENCES

Available upon request.