Dong Dai

630 S 4th St, Unit 30, Ames, IA, 50010 | Tel: (515) 708-4177

Email: daidong1222@gmail.com | Github: https://github.com/dongdai23

EDUCATION

- Ph.D. Statistics and Applied Mathematics (Co-major), **Iowa State University**, Ames, IA   08/2013-05/2019
- B.S. Mathematics, **Sichuan University**, Chengdu, Sichuan, China   09/2009-07/2013

TECHNICAL STRENGTHS

Programming Skills   R, Python, SQL, Tableau, SAS, MATLAB

Related Experience   Data Visualization, Data Mining, Machine Learning, Deep Learning, Generalized Linear Model, Bayesian Analysis, Survival Analysis, Time Series, Optimization, Experimental Design

PROFESSIONAL EXPERIENCE

Department of Mathematics, ISU, **Teaching Assistant**   Ames, IA

Teaching Lectures and Grading   08/2013-Present

- Lead recitation lectures of undergraduate math classes, such as Calculus, Ordinary Differential Equation.
- Teach stand along Pre-Calculus course of 58 students, design course syllabus and exams for the course.
- Interact with students during office hours to answer their questions regarding course material.

Home Credit US, Risk Modeling Intern   Overland Park, KS

Data Visualization, Data Science and Predictive Modeling   06/2018-08/2018

- Built models of imbalanced data to rank-order and separate good customers from bad customers and evaluated models based on Receiver Operating Characteristic (ROC) curve.
- Compared coverage of different data resources to help cost-benefit analysis using Venn diagram along with statistics.
- Developed an automated reporting system using histograms and summary statistics of multi-dimension by R and Python.
- Visualized the relation of important features to the predictor using linear regression by Tableau and Python (Matplotlib).
- Verified the consistency of data resources using an automated way by visualization as well as metrics of description.

RESEARCH AND PROJECT EXPERIENCE

Daily Gasoline Production Prediction   Ames, IA

Analyzed time series data using ARIMA and Deep Learning models   11/2018-01/2019

- Aggregated and visualized data from different files using Python (matplotlib, ploty).
- Utilized Autoregressive Integrated Moving Average (ARIMA) model and Long Short Term Memory (LSTM) Recurrent Neural Network (Tensorflow, Keras) to predict daily gasoline production.
- Tuned parameters of Recurrent Neural Network models to reduce 21% error comparing to the naïve model.
- Compared predictions to the actual gasoline production by visualization.

Cost-Benefit Analysis of Medication Assistant   Ames, IA

Built models to analyze whether the cost for assistant worth the benefit to patients (health care)   04/2018-12/2018

- Extracted the data from sample survey and applied data visualization for different attributes.
- Applied Poisson regression to predict the number of medication errors made by patients who forgot or omitted the previous medication taken when meeting doctors.
- Built generalized linear model for time consumed by the pharmacy technician to help verify medication taken by patients.
- Conducted cost-benefit analysis and interpreted the results to managements.

Kaggle Competition of House Prediction, Obtained Top 6%   Ames, IA

Predicted House Prices utilizing data Mining and Machine Learning tools and methods   04/2017-07/2017

- Identified the key component and attributes using data visualizations and analysis.
- Conducted data preprocessing, including filling missing values with different methods (median or create dummies) and feature engineering by adding economic data (Numpy, Pandas).
- Trained machine learning models such as Random Forest, XGBoost, Elastic Net, Neural Networks, etc. using Python (Scikit-learn, Keras, Tensorflow).
- Tuned parameters of models based on cross-validation loss, feature importance and training curve.
- Created final model by stacking different models depending on their cross-validation loss, which decreased 10% final error.

Israel's Bank Call Center Data Analysis   Ames, IA

Simulated queue processes using real world data and different staffing methods   01/2016-8/2017

- Applied data visualization using R (ggplot2) to obtain the distribution of customers' call in, unhold, servicing time.
- Applied Kaplan-Meier estimator to analyze the actual waiting time (censored data) distribution.
- Compared simulation results of different staffing rules based on visualizations of key statistics.
- Proved convergence of iterative staffing algorithm, which performed best in the simulation results.