

JIEYI DENG

 <https://github.com/Jieyi-Deng> |  jieyi.deng@rutgers.edu |  (848)-252-8487 |  www.linkedin.com/in/jieyi-deng-

SKILLS

- Programming Skills: Python, SAS, R, MySQL, Matlab, PySpark, Scikit-Learn, TensorFlow
- Industrial Knowledge: Machine Learning Modeling, Data Analysis, ETL Process, Natural Language Processing (NLP), Hadoop, Exploratory Data Analysis (EDA), Data Visualization, Hypothesis Testing

Professional Experience

GEICO **Washington, D.C**

Modeling Programmer (Data Scientist)

03/2019 – present

- Responsible for data mining and exploration in the department of Claim Home Office (CHO) and support various research projects with extensive data analysis and machine learning model application.
- Developed customer satisfaction prediction based on valid post-FNOL (First Notice of Loss) surveys from the past 2-year. This project testifies the expectations toward to adjusters are most important to the overall satisfaction.
- Tracked hurricane paths from the past to model the risk for each catastrophic hurricane and simulated the advisory tracks within 5 days for the oncoming hurricane to estimate the number of potential claims before the storm landfall.

Waldron Inc.

New York City, NY

Data Scientist (Intern)

01/2019 – 03/2019

- Designed and developed the reinforcement learning system to offer the basic recommendation system for an online shopping APP, the major tasks include fake data construction, feature engineering and reward/regret strategy application.

Center for Advanced Infrastructure and Transportation (CAIT)

New Brunswick, NJ

Data Scientist

09/2016 – 09/2018

- Engaged in data extract, transform and load (ETL) process based on SQL to refine engineering problems; employed feature extraction and visualization to provide data explanation and in-depth comprehension.
- Performed statistical models or statistical testing on R to measure the risk of stochastic events and explained the relationship between accidents and potential accident-causing factors.
- Trained various predictive models based on Jupyter Notebook (e.g. logistic regression, ensemble methods) to predict the frequency of events and in charge of cross-validating results and unifying the solution among team peers.

Projects

Risk Analysis and Prediction of North American Rail Accidents 08/2017 – 09/2018

- Recognized pattern and predicted frequency for railway accidents based on historical records, e.g. maintenance activities.
- Collected and combined the pieces of data from various departments and preprocessed it with ETL approach. Then applied EDA to understand the data and find scenarios for further performing the analysis.
- Used Logistic Regression model as a benchmark and validated model results by Conditional Odds Ratios (COR) to further explain the effects of causing factors on the stochastic accident rates.
- Extracted 30 engineering features and built predictive models using Random Forest, XGBoost and Feedforward Neural Network. The corresponding Receiver Operating Characteristic (ROC) curves are 0.78, 0.87 and 0.85 with 5-fold cross-validation.

Probabilistic Risk Analysis of Flying Ballast Hazard on the High-Speed-Rail (HSR) Lines

09/2016 – 08/2017

- Statistical modeling and risk simulation for a stochastic event in High-Speed-Rail engineering with limited experimental data.
- Modeled the occurrence of the event as a balanced system and extracted the process into discrete mechanical factors (e.g. wind force, gravity) and mathematically connected each factor with common operating parameters (e.g. train speed, particle density).
- Collected operating data and records from in-situ experiments and normalized the statistical distribution of each mechanical factor, then deduced the risk as a cumulative probability.
- Integrated the analyzing process to a flexible Probabilistic Risk Analysis framework and evaluate how the risk increases as 0.019% to 1.2% with the train speeding from 250 km/h to 400 km/h.

Natural Language Processing and Pet Owner Classification on Video Comments

08/2018

- Identified the pet owners from the video comments and explored the significant topics for them.
- Preprocessed the comments by removing invalid text and unknown creators, and identified the comments with the label of pet owners or not.
- Extracted features by using tokenizer to split each comment; employed Word2Vec to map each word to a unique fixed-size vector and then transformed each word into a vector using the average of all words in the comment.
- Implemented a Logistic Regression classification to identify the pet owners from comments with regularization parameter grid searching, the area under the ROC curve is 0.91.
- Defined a text analysis function to categorize significant topics to the pet owners by Latent Dirichlet Allocation (LDA) after removing stop words and special characters from the text and the most significant topics are related to training, food and life span.

PUBLICATIONS

- [Probabilistic Risk Analysis of Flying Ballast Hazard on High-speed Rail Lines](#), 2018, Transportation Research Part C
- **A Literature Review of Rail Defect Causal Factors**. In 2018 Joint Rail Conference, American Society of Mechanical Engineers.
- **Probabilistic Analysis of Flying Ballast on High-Speed Rail Lines**, In 2017 Joint Rail Conference, American Society of Mechanical Engineers.