SCC Seminar Series on Data Analysis

Weifeng Wang
Consultant, UGA Statistical Consulting Center

Tuesday, November 12, 2019
12:30~1:30pm in room 102, Caldwell Building

My Experience as a Data Science Intern at Lincoln Financial Group

The data scientist role is becoming more and more important as businesses rely more heavily on data analytics to make decisions.

This summer I worked as a Data Science Intern at Lincoln Financial Group to assist colleagues on medical/business research in a hands-on environment. My main job was to apply statistical models and ML/DL algorithms on unstructured data including text and voice from Lincoln's Life Solutions Underwriting process to predict impairments includ ing cancers of insureds.

This presentation will cover one of the projects I worked on, which is the impairment classification of underwriting notes with text mining. I will talk about how to conduct text preprocessing and feature extraction with bag-of-words & tf-idf models, how to deal with multi-labeled samples with user-defined filter dictionary, how to deal with imbalanced data in text format with up-sampling method by adding noisy words and down-sampling method with a filter dictionary.

In this project, we built a two-step classifier model based on xgboost algorithm and got precision 92%, recall 93% and F-score 91% in testing data (10,000 insureds). Also, our model successfully detected mislabeled notes in 2019. Finally, we implemented our model to the unlabeled samples of 580,938 insureds from 2012 to 2018.

Finally, I will talk about what I learned from this summer intern and what skills are necessary if you want to work as a data scientist or data analyst in a financial services firm like Lincoln Financial Group.