(805)-744-5022 · coronakhalel@gmail.com · https://www.linkedin.com/in/khalel-corona

Dear Recruiting Department,

In the process of receiving my degree, I got to be research assistant for the BYU economics department, enabling me to hone my capacity to leverage data science for real-world needs. I assisted the BYU economic research department in automating genealogical work through machine learning, using a classification algorithm to assign new individuals in historical documents as a match to ancestral family trees, pedigrees. I have done genealogical work for many years, so my role in this project was to use my specialized knowledge to verify that the algorithm's predictions were at least as good as my own predictions. Ultimately, I verified several thousands of predictions, helping our algorithm reach over 90% accuracy. Though my job wasn't necessarily computationally intensive, I received the opportunity to get up close and personal with the foundations of machine learning, producing quality data from which an ML method can derive reliable test accuracy. This project was performed for the business FamilySearch through our department and has led to an unbelievable exponential growth in pedigree connection for under-represented peoples. For example, thanks to our algorithm, African-Americans now have pedigrees that are even more connected than even white Americans. I have also conducted my own projects for the BYU economics department. In one of my personal research projects, I quantified the effect of implementing a US state mask mandate on covid case growth rates. Amazingly, through quantile regression analysis and difference-in-differences regression analysis, I was able to observe that in the several months that followed a state's mask mandate implementation, there was a general decay in the growth of covid case rates. Further interested in covid data, I conducted another research project for our department using machine learning. In this project, my team aimed to predict the excess mortality of the US in 2020 caused by covid, drawing from hundreds of mortality predictors and thousands of US county-level data. Using python to leverage modules, pandas to construct dataframes, numpy to clean-up datasets, and scikit-learn to train ML models (penalized regression, gradient boosted regression, and random forests) we achieved results very similar to the CDC, The Economist, and the World Health Organization concerning the US's 2020 excess mortality. My team used some beautiful data visualizations that depict the results of our project in a way that is far superior to linguistic description. I invite you to take a look for yourself through this hyper-link to my small portfolio. Both of these personal data projects consisted of compiling datasets that entailed hundreds of thousands of observations, sound data preprocessing, and intuitive statistical modeling. Currently, I am focused on refining my ability to leverage neural networks using more advanced Python modules such as Tensorflow and Keras. I do believe that I am grasping the theory and coding just as well as I have grasped the methods available in Scikit-learn's ML library. Through implementation, practice, and education, I believe that I have acquired an extremely deep familiarity with the statistical theory and coding required for data science. I thank you for your time in considering me. I sincerely hope to be selected for an interview so that I can personally express my interest in this position. Please don't hesitate to contact me with any questions!

Sincerely,