Research Statement

Farid Razzak (farid.razzak@rutgers.edu)
Rutgers Business School

My general areas of research have been data mining & business analytics with applications in financial services, market research, & regulations. Specifically, the focus of my research currently is to develop efficient & effective techniques to address various *Bank Stress Test Analytics* challenges, to discover & understand the relationship amongst exogenous economic factors & banking performance profiles to intelligently provide scenario-based banking capital forecasts. My work has been published in top data mining conference **IEEE ICDM** with extensions of the work to be submitted to relevant academic journals & conferences. In the remainder of this statement, I will further describe several specifics of the research with focus on multi-modal generative economic forecasting, scenario-based banking capital & loss prediction, transferable knowledge of financial banking systems, & conclude with an overview of future directions.

Economic Conditions Estimation

In top-down methodology approaches to bank stress testing, exogenous economic factors are believed to have influence on the trajectory of a bank's performance. Therefore, considering factors that are commonly used to understand the overall state of global, national, industrial, or financial health are crucial to understanding how banks are likely to be affected by economic conditions. Economic forecasting methods typically focus on domestic & international macro variables when depicting conditions for financial systems. While most existing studies focus on structural linear approaches for forecasting, little focus is given on finding relationships among both macro & micro economic variables that may depict the financial economy in a more appropriate manner.

To this end, this work focuses on providing dependent non-linear joint probability distributions among the exogenous factors while considering data availability & accessibility constraints. Along this line, I first identify additional macro & micro economic variables beyond that of standard regulatory practices to help provide additional granularity to the economic analysis. Then, we sample estimation from a multi-modal data distribution model trained on historical economic records to provide the most likely estimations. The implementation of multimodal conditional generative model that first extracts the shared latent features to then learn the dependent joint representations conditioned on a single input modality while considering multi-modality observation & estimation challenges, is used to robustly perform economic forecasting. The proposed framework utilizes a **Multimodality Conditioned Variation Autoencoder model (MCVAE)** to achieve such a mechanism. This model aims to estimate the desired economic variables of different aspects given one existing variable. Consisting of two components, the encoding & decoding networks, the model first passes multiple modalities & a separate conditional modality through an encoding network to learn hidden representations of shared variables across modalities. After that, given the conditional modality & any initial variables sampled from Gaussian distribution, the decoding network can finally produce the estimation of target modalities. In our problem task, the target input modalities & conditional input are different aspects of the economic conditions while the goal of MCVAE is to generate or simulate the overall economic conditions when given only one specific type of economic condition. With considerations for multi-modality estimation, data availability & accessibility, the proposed framework is able to improve overall performance in estimating real-world economic conditions when compared to baseline methods. This framework is key to generating likely economic estimations or simulated adverse economic conditions for the purposes of scenario-based banking performance prediction. A summary of the preliminary work for this framework has

been published in **IEEE Conference on Data Mining (ICDM'19)** as the first of two stages of an integrated framework for banking stress test analytics. Extended versions of this work with additional experimentation & case studies is currently being worked on to be submitted to an appropriate journal.

Banking Capital & Loss Prediction

Another aspect of the top-down stress test methodology is the reliance on publicly released bank financial statements for the assessment of the respective bank's loan portfolio, loan loss rate, & net revenue, which represent significant factors related to projections of capital adequacy under different economic conditions. Considerations of the dynamic nature of dimensional influence on prediction over a temporal space are needed to properly capture differing financial crisis. Specifically, different banking & economic conditions factors would influence banking performance at different time periods, which can make generalization more complex. In our work, there are two types of features that can be used to predict future bank capital & loss. 1) Exogenous factors that mainly consists of economic conditions, from micro to macro economic perspective; & 2) Historical bank performance values of the respective time-series. These two factors are integrated in our work by applying a **Dual-Stage Attention-based neural network model (DA-RNN)**, which takes both attention mechanism & long short-term temporal dependencies (LSTM) into consideration for better time-series prediction. Specifically, DA-RNN consists of two LSTM networks that incorporate attention mechanism to select relevant features. The dimensional & temporal consideration of DA-RNN model makes it an effective candidate for our prediction task given that different banking performance profile & economic condition characteristics may have influences on the capital & loss ratios at different time periods. This tends to be true for the financial industry, given recent crisis' (e.g., housing crisis, internet dotcom bubble, etc) were predominately influenced by different banking & economic factors. Through extensive experimentation on banking capital & loan loss ratio predictions from real-world U.S. banking data , we are able to determine the utility of the proposed framework for the purposes of banking stress test analytics. A summary of the preliminary work has been published in **IEEE Conference on Data Mining (ICDM'19)** as the second part of an integrated framework that combines economic conditions estimations with bank capital & loss prediction.

Other Achievements

I have also worked on various topics & professional applications related to business analytics & data mining, including indoor occupancy analytics, predictive autoscaling for cloud workloads, & trade surveillance detection algorithms. For example, during my summer internship at *Philips Research*, I performance extensive analytics & developed dashboards for a prototypical occupancy sensor data to help optimize lighting, cleaning, & security costs for large commercial buildings. During my summer research internship at **IBM T.J. Watson**, I conducted experiments on a private cloud cluster where I coordinated stress workloads on cloud instances to record performance data to then conduct time-series classification & clustering on patterns that could help auto-scaling features preemptively deploy additional resources for the purposes of competitive advantage against other cloud management providers. During my graduate internship at the **Securities & Exchange Commission's Quantitative Analytics Unit**, I developed analytical tools & bespoke features to help analyze for federal securities law violations from financial institutions. Most recently, during my time as a Data Scientist at **ACA Decryptex Financial Laboratories**, I helped develop & test a financial trade surveillance software product for financial institutions to analyze their end of day trade transactions & order-flow messages for potential federal securities law violations based on advanced analytical techniques. After the development of the product, I attended various

professional discussion panels, conferences, webcasts, & onsite presentations to discuss technical & analytical aspects of the software. I have also published an informative chapter title "Climate Risk Analysis for Financial Institutions" in **Enyclopedia of GIS**, which touches on climate finance & how financial institutions utilize geographic information systems (GIS) for investment research.

Future Research Direction

It is my belief that data-driven solutions to business challenges will continue to evolve as technology & state-of-the-art research grows. My current & future research ambitions seeks to focus on the intersection of financial services, technology, & advanced analytics to address both research & business challenges.

In the near future, I hope to continue my research in applying advanced analytical techniques to address research opportunities in the fields of FinTech & RegTech using the latest technological tools available. Specifically, I hope to continue in-depth research in Bank Stress Test Analytics & contribute more to the literature of economic forecasting, balance sheet projections, & scenario-based capital adequacy. Traditionally, the research in these areas focus mainly around statistically linear & structural approaches, while I would focus on applying robust deep learning methodologies, non-linear techniques, data mining, & machine learning approaches to leverage the rich but unstructured heterogeneous financial data that is available.

Particularly, the work using deep learning techniques to generate economic conditions estimates to then use in a dual-stage recurrent neural network for the purposes of balance sheet capital & loss ratio predictions in an effective manner is one of very few attempts to address this task using sophisticated analytics.

Currently, I have working papers that will explore aspects of transfer learning techniques that can be leveraged while considering positive/negative transfer to find underlying commonalities among financial banking systems that transcend geo-political, geo-economical & geo-regulatory impacts. Using similar balance sheet features from US banks & Chinese banking sector to first pre-train a model based on aggregated similarity based features to address negative transfer aspects of the source domain (US banks). Then, transfer model parameters to finally train based on the target domain (Chinese Bank) to understand if the model can build on fundamental commonalities learned from the source to create more specialized parameters yielding better prediction performance when compared to predictions without pre-trained aspects. I plan to test this framework against European banks & other banking regions to establish transferable knowledge from each region that could be beneficial & shine light on universal banking behaviors. I plan on completing the experimentation on this work shortly & submitting it to a top data mining conference.

I am also co-authoring work on social media trend popularity prediction research using sentiment analysis, natural language processing & deep learning methodologies to provide potential hashtags for digital marketing campaigns specific to industry based on competitor campaigns, trending hashtags, & historical analysis of social media impression influence metrics. I hope to use techniques developed from this work to be applied to more financial domain specific applications pertaining to investment research & regulations.