CSBS 2023 ANNUAL DATA ANALYTICS COMPETITION
RESEARCH PROPOSAL

1 University/College Name:

Georgetown University, 3700 O St NW, Washington, DC 20057

2 Participant List:

Name	Email	Graduation Year	Major
Yihan Bian	yb214@georgetown.edu	2024	Data Science and Analytics
Qiaojuan Tu	qt34@georgetown.edu	2024	Data Science and Analytics
Yunhan Zhang	yz1048@georgetown.edu	2024	Data Science and Analytics

3 Faculty Sponsor:

Dr. Nakul R. Padalkar (np617@georgetown.edu),

4 Course:

ANLY 512: Statistical Learning for Analytics

5 Research Idea/Hypothesis:

The United States is entering into the economic recovery from the COVID-19 crisis in roughly two years. With it, the incremental need to curb inflation dominates the monetary policy discussion. Resulting of the fiscal policy during covid time dictated by the past and current administrations, an increasing clamor for the Federal Reserve to raise interest rates to relieve inflationary pressures is gaining momentum.

The Fed has repeatedly increased Federal Funds Rate since March 2022, the interest rate for overnight borrowing between banks, to help cool down the high inflation. To better understand the behavior of the banking industry during a rapidly rising interest rate environment, it is crucial to examine historical interest rate hikes and how different banks navigated them. For this reason, we focus on comparing the survival probability of the banks that survived under the 2004-2007 hike scenario in recent 2022.

6 Expected Data Source:

1. FFIEC Call Reports Data Reporting Year 2004-2007 and 2020-2022
2. FFIEC CDR UBPR Ratios Summary Ratios 2004-2007 and 2020-2022
3. FRED Federal Funds Effective Rate
4. Federal Deposit Insurance Corp.

7 Proposed Methods to Test Hypothesis:

Our hypothesis posits that the statistical characteristics of banks will exhibit a comparable alteration as a result of the increment in interest rates. To test this hypothesis, we intend to implement a classification model utilizing the Uniform Bank Performance Report (UBPR) statistics obtained from the Federal Financial Institutions Examination Council (FFIEC) as inputs. The label for this model will be the change in the scale of each bank. Through this model, we aim to predict the present scenario and determine the proportion of vulnerable banks to insolvency.

In order to achieve our objective, we will first acquire and prepare the requisite UBPR data from FFIEC. This will entail cleaning and organizing the data to ensure it is in an appropriate format for modeling. The key data we are seeking is from the period around 2004-2007, when the Federal Reserve decided to increase the interest rate. Subsequently, we will execute and train our classification model utilizing a suitable algorithm, such as Logistic Regression, Support Vector Machine, or a Multi-Layer Perceptron (for Prediction).

After the model has been trained, we will assess its efficacy using accuracy, precision, and recall metrics. We will then utilize it to make predictions on the present scenario. Additionally, we will utilize the model to determine the proportion of banks vulnerable to insolvency and supply recommendations for reducing this vulnerability.

The methodology will offer a comprehensive and orderly method for determining the proportion of banks at risk of bankruptcy by analyzing their UBPR statistics and the alteration of interest rates, as described in the previous paragraphs.

8 Planned Deliverables:

We will use the business intelligence dashboard (BI dashboard) to interactively show two different data visualization solutions. First, we will use an interactive slider and scatterplot to show the change in the size of all banks as the interest rate changes. Second, we will draw an abstract bank as a standard based on the prediction model mentioned above. Then compare the input data of the actual bank with this standard to demonstrate which indicators of this existing bank are in the danger zone or safe zone and make recommendations based on this information.