Conference of State Bank Supervisors

The Role of Community Banks During the Pandemic

CSBS DATA ANALYTICS COMPETITION

MELISSA SERRANO
ANNALISE SUMPON
MADI TEDROW

ADVISOR: ELI OLINICK

Southern Methodist University

SMU

Our Team

Madi Tedrow

Management Science/
Economics
Global Markets Analyst -
Goldman Sachs

Melissa Serrano

Management Science/
Data Science
Global Consumer Technology
Analyst - Citi Bank

Annalise Sumpon

Management Science/Data
Science/Mathematics
Data Scientist - IBM

SMU

Agenda

Problem and background

- Problem
- Goal
- PPP Loans
- Legislation relevancy
- Stakeholder requirements

Data preparation

- Data cleaning
- Variable selection
- Model selection

Modeling

- Logistic regression
- CART model
- Naive Bayes

Results

- Model Comparison
- Goal Reflection

Conclusion

The Problem

The Small Business Administration (SBA) set up the Paycheck Protection Program (PPP) loan program in response to the pandemic.

Preliminary research conducted by the Conference of State Bank Supervisors (CSBS) shows that community banks were at the forefront in providing these needed funds to small businesses

Are specific communities or business characteristics affiliated with the role of community banks in the distribution of PPP loans?

Our Goal

Determine if there exists statistically significant differences or patterns in the **business type**, **business size** (number of jobs), and **business ownership** (e.g., racial minority or gender) for the businesses that received the PPP loans from community banks and/or non-community banks.

Why is this important?

- PPP Loans
- The role of community banks
- FDIC designation label
- Opportunities in legislation

PPP Loans

Paycheck Protection Program (PPP)

- Loans provided to small businesses during the COVID-19 pandemic to assist businesses keep their workforce employed during the pandemic
- Many of the PPP loans are being/have been forgiven

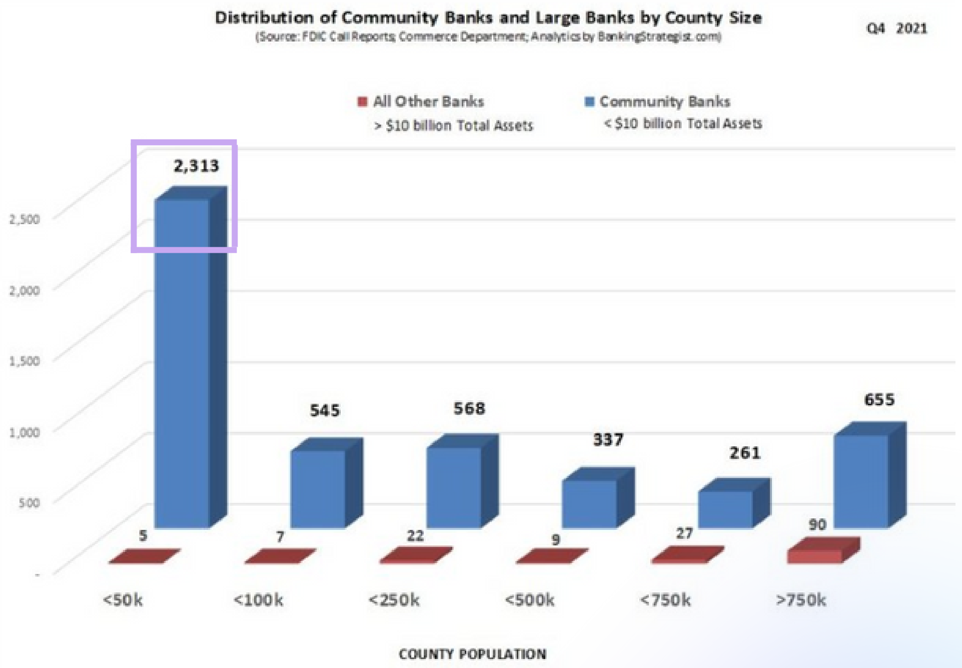
PPP Controversies

- 23% - 34% of PPP dollars were used towards directly helping workers
- Publicized concerns about who received the loans arose after the distribution of the PPP loans.

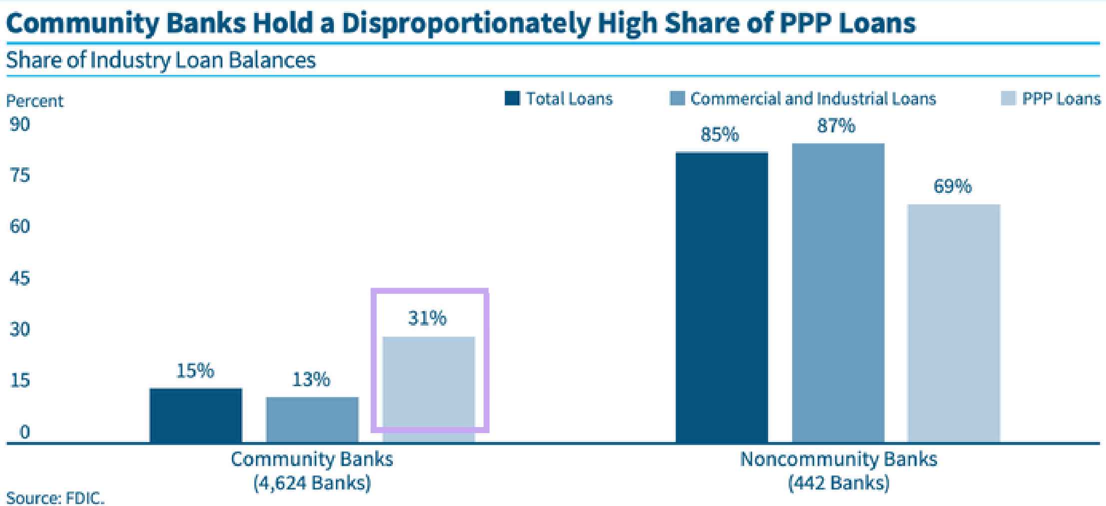
- "Did Asian- and American Indian-Owned Businesses Receive Their Fair Share of PPP Loans?" *(Heartland Forward - August 11, 2020)*
- "The Paycheck Protection Program Failed Many Black-owned Businesses" *(Vox - October 5, 2020)*
- "Minority Entrepreneurs Struggled to Get Small-Business Relief Loans" *(New York Times - April 4, 2021)*
- "Racial Bias Affected Black-owned Small Businesses Seeking Pandemic Relief Loans, Study Finds" *(The Washington Post - October 15, 2021)*

The Role of Community Banks

- Serve a critical role in smaller and rural markets across the US by providing credit, liquidity, and investments to communities
- Typically in counties with populations of 50,000 or less



The Role of Community Banks



Source: FDIC.
Note: Not all banks in the count displayed hold Paycheck Protection Program (PPP) loans: 83 percent of the 4,624 community banks and 73 percent of the 442 noncommunity banks hold PPP loans. Data as of June 30, 2020.

- Distributed Paycheck Protection Program (PPP)1 loans during the Covid-19 pandemic
- Held $148 billion—28 percent of total PPP loans and 31 percent of PPP loans held by banks

FDIC Designation

5 STEP PROTOCOL FOR DESIGNATION

Aggregation, filtering, inclusion, scope identification, asset size determination

Community banks are successful in areas with growing economies and populations while they continued to meet the credit needs of less economically vibrant areas, such as rural counties and areas with declining populations.

Need FDIC designation to be formally recognized as a community bank and thus be correctly classified receive a sufficient amount of support

"The FDIC recognizes the critical role community banks play in providing loan and deposit services to customers throughout the United States. By continuing to study community banks, the FDIC can provide support to these institutions and the communities they serve."

FDIC Chairwoman Jelena McWilliams

Opportunities in Legislation

STRENGTHEN	IDENTIFY	EXPAND
Support communities and counties that community banks serve with lower income and minority populations.	Create special designation labels for banks with community bank characteristics.	Expand credit willingness for community banks to receive loans as prevention against financial downwind.

Data Preparation

Data Cleaning

Cleaned merged datasets by extracting needed columns and filling or dropping null values

Variable Selection

Constructed statistical tests to identify early key variables and remove highly correlated variables

Model Selection

Considered a variety of models to identify key variables to predict community banks from non-community banks

Data Cleaning

- Merged the provided CSBS dataset of with the original SBA "PPP plus 150k" dataset
  - Purpose: to combine additional variables that may provide insight into our goal

- Extracted needed columns and filled or dropped null values
  - Purpose: to avoid duplication and specify variable scope

CSBS Dataset
8,614,374 observations with 19 variables

Merged Dataset
8,614,374 observations with 58 variables

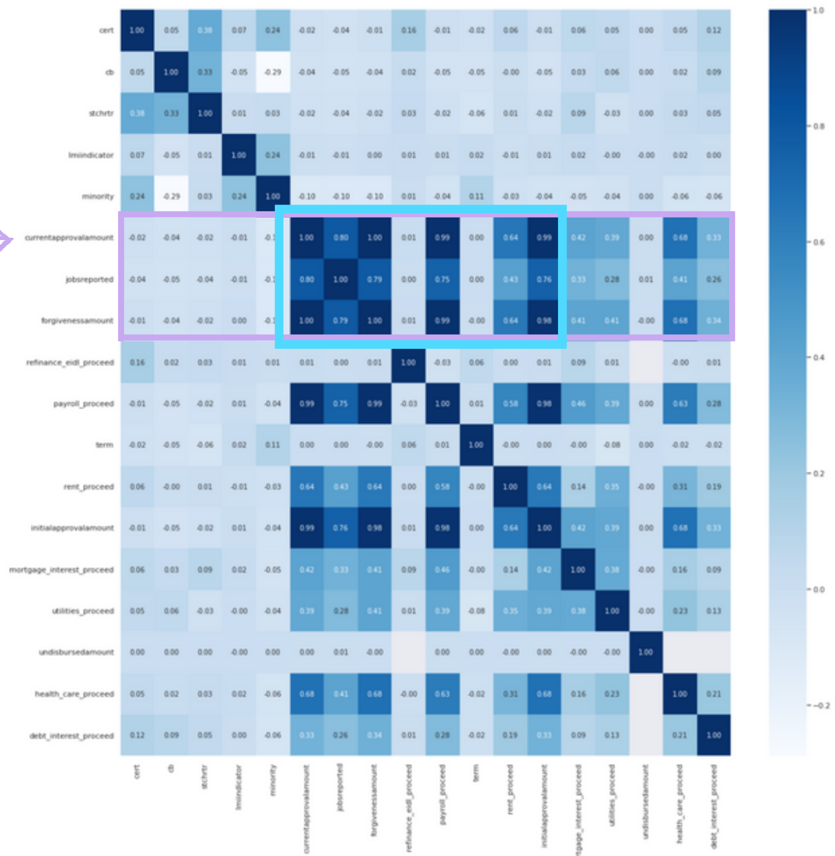
The resulting dataset: 15 variables with 122,155 observations

Variable Exploration

Correlation Matrix

High correlation with *currentapprovalamount*

- initialapprovalamount
- forgivenessamount
- jobsreported
- payroll_proceed



Variable Selection

Chi-Squared Test

$$\chi = \sum \frac{(Observed - Expected)^2}{Expected}, \ Expected = \frac{RowTotal \times ColumnTotal}{OverallTotal}$$

H_o: *No association with Community Banks*

H_a: *There is evidence of association with Community Banks*

Features	F Score	P values
stchrtr	784,199.77	0.000E+00
ruralurbanindicator	161,810.14	0.000E+00
minority	123,326.14	0.000E+00
lmiindicator	44,557.45	0.000E+00
race White	5,648.44	0.000E+00
businesstype_Corporation	1,915.60	0.000E+00
businesstype_Limited Liability Company(LLC)	1,210.14	3.825E-265
businesstype Subchapter S Corporation	1,206.07	2.923E-264
businesstype_Non-ProfitOrganization	736.64	3.218E-162
race _American Indian or Alaska Native	246.22	1.733E-55

Mutual Information Selection

$$MI(A,B) = \sum_x \sum_b p(x,b) \times log(\frac{p(x,y)}{p(x)p(y}) \leq 1$$

Features	Mutual Information
currentapprovalamount	0.006314
payroll_proceed	0.005663
health_care_proceed	0.004669
utilities_proceed	0.003690
mortgage_interest_proceed	0.002695
forgivenessamount	0.002352
debt_interest_proceed	0.000213
refinance_eidl_proceed	0.000154
jobsreported	0.000000
rent_proceed	0.000000

Final Variable Selection

cb	0 = Non-Community Bank, 1 = Community Bank
stchrtr	0 = Non-State Chartered , 1 = State Chartered
ruralurbanindicator	U = Urban, R = Rural
lmiindicator	0 = Not in low-to-moderate income area, 1 = In Low-to-moderate income area
currentapprovalamount	Loan amount approved
jobsreported	Number of jobs the business has reported
forgivenessamount	Loan amount forgiven
race	Borrower Race Description
minority	0 = No minority owned business , 1 = Yes minority owned business
ethnicity	'Unknown/NotStated', 'Not Hispanic or Latino', 'Hispanic or Latino'
businesstype	Business Type Description
payroll_proceed	Amount of proceeds that were reported to be applied towards Payroll costs

Model Selection

Linear Regression	Polynomial Regression with Multiple Variables	Extreme Gradient Boosting	K-means Clustering
Logistic Regression	K-nearest Neighbors	Classification and Regression Trees	Naive Bayes

Modeling

- Logistic Regression
- Classification and Regression Trees (CART)
- Naive Bayes

Logistic Regression

Purpose to use categorical inputs to describe create and purposeful logistic equations for classification

Initial Model Performance

```
             Logistic Regression Initial Results
-----------------------------------------------------------
**************** Evaluation on Test Data ****************
Accuracy Score 0.6864409092149426

              precision    recall  f1-score   support

           0       0.60      0.72      0.66     15243
           1       0.77      0.66      0.71     21404

    accuracy                           0.69     36647
   macro avg       0.69      0.69      0.68     36647
weighted avg       0.70      0.69      0.69     36647

-----------------------------------------------------------
```

cb, stchrtr, ruralurbanindicator, lmiindicator, currentapprovalamount, jobsreported, forgivenessamount, race, businesstype, payroll_proceed, and minority

Importance Coefficients

Top 15 Coefficients for Logistic Regression Initial

	coef
stchrtr	6.111765
race_American Indian or Alaska Native	3.711375
businesstype_Housing Co-op	2.343501
race_Puerto Rican	2.278674
ruralurbanindicator_U	1.688637
businesstype_501(c)6 – Non Profit Membership	1.652083
businesstype_Rollover as Business Start-Ups (ROB	1.604697
businesstype_Tenant in Common	1.571997
payroll_proceed_large	1.549992
businesstype_Self-Employed Individuals	1.524092
businesstype_Independent Contractors	1.490503
businesstype_Partnership	1.418378
businesstype_Non-Profit Childcare Center	1.371255
businesstype_Limited Liability Company(LLC)	1.347864
businesstype_Corporation	1.343657

Logistic Regression

Final Model Performance

```
                Logistic Regression Final Results
-----------------------------------------------------------
**************** Evaluation on Test Data ****************

Accuracy Score 0.687777990012825

              precision    recall  f1-score   support

           0       0.60      0.74      0.66     15243
           1       0.78      0.65      0.71     21404

    accuracy                           0.69     36647
   macro avg       0.69      0.70      0.69     36647
weighted avg       0.71      0.69      0.69     36647

-----------------------------------------------------------
```

cb, stchrtr, ruralurbanindicator, lmiindicator, currentapprovalamount, race, businesstype, and payroll_proceed

Importance Coefficients

```
Top 15 Coefficients for Logistic Regression Final
                                                      coef
stchrtr                                           6.122666
race_American Indian or Alaska Native             3.860552
race_Puerto Rican                                 2.346440
businesstype_Housing Co-op                        2.339864
ruralurbanindicator_U                             1.696062
businesstype_501(c)6 – Non Profit Membership      1.673439
businesstype_Rollover as Business Start-Ups (ROB  1.579138
businesstype_Self-Employed Individuals            1.555948
businesstype_Tenant in Common                     1.534472
businesstype_Independent Contractors              1.507943
businesstype_Partnership                          1.419181
businesstype_Non-Profit Childcare Center          1.357319
businesstype_Limited  Liability Company(LLC)      1.343868
businesstype_Corporation                          1.339409
payroll_proceed_large                             1.292922
```

Classification and Regression Trees (CART)

Calculate Gini Impurity scores to predict class and numeric labels for community bank loans

Initial Model Performance

```
*************** Tree Summary ***************
Classes:  [0 1]
Tree Depth:  3
No. of leaves:  8
No. of features:  9
--------------------------------------------------------

*************** Evaluation on Test Data ***************
Accuracy Score:  0.6855859000087252
              precision    recall  f1-score   support

           0       0.76      0.66      0.71     13259
           1       0.61      0.72      0.66      9663

    accuracy                           0.69     22922
   macro avg       0.68      0.69      0.68     22922
weighted avg       0.70      0.69      0.69     22922
```

stchrtr, ruralurbanindicator, payroll_proceed, minority, jobsreported, forgivenessamount, currentapprovalamount

Feature Importance

Feature Number	*Variables*	*Feature Importance Score*
0	minority	0.0691
1	stchrtr	0.7437
2	ruralurbanindicator	0.1849
3	jobsreported	~0
4	currentapprovalamount	~0
5	forgivenessamount	~0
6	lmiindicator	~0
7	businesstype	~0
8	payroll_proceed	0.0023

Classification and Regression Trees (CART)

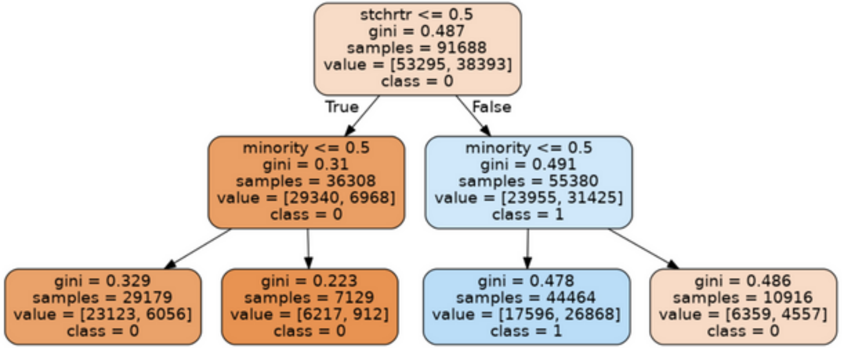
Calculate Gini Impurity scores to predict class and numeric labels for community bank loans

Final Model Performance

```
*************** Tree Summary ***************
Classes:  [0 1]
Tree Depth:  2
No. of leaves:  4
No. of features:  2
--------------------------------------------------------

*************** Evaluation on Test Data ***************
Accuracy Score:  0.6890759968589129
              precision    recall  f1-score   support

           0       0.76      0.68      0.72     13259
           1       0.61      0.70      0.66      9663

    accuracy                           0.69     22922
   macro avg       0.69      0.69      0.69     22922
weighted avg       0.70      0.69      0.69     22922

--------------------------------------------------------
```

stchrtr, minority

CART Reduced Tree Graph



Naive Bayes

Assumes variable independence and makes use of probabilities to make predictions

Bayes' Law:

$$P(C \mid A) = \frac{P(A \cap C)}{P(A)} = \frac{P(A \mid C)P(C)}{P(A)}$$

CSBS Application:

$$P(cb \mid features) = \frac{P(features \mid cb)P(cb)}{P(features)}$$

Naive Bayes

Assumes variable independence and makes use of probabilities to make predictions

Initial Model Performance

```
              Naive Bayes Results
------------------------------------------------------------
***************** Evaluation on Test Data *****************

Accuracy Score 0.6839426460750953

              precision    recall  f1-score   support

           0       0.77      0.65      0.70     19853
           1       0.60      0.73      0.66     14530

    accuracy                           0.68     34383
   macro avg       0.69      0.69      0.68     34383
weighted avg       0.70      0.68      0.69     34383

------------------------------------------------------------
Confusion Matrix:
[[12922  3936]
 [ 6931 10594]]
```

borrowerstate, stchrtr, ruralurbanindicator, lmiindicator, currentapprovalamount, forgivenessamount, minority, race, businesstype, and payroll_proceed

Final Model Performance

```
              Naive Bayes Results
------------------------------------------------------------
***************** Evaluation on Test Data *****************

Accuracy Score 0.63572114126167

              precision    recall  f1-score   support

           0       0.63      0.90      0.74     19853
           1       0.67      0.27      0.39     14530

    accuracy                           0.64     34383
   macro avg       0.65      0.59      0.56     34383
weighted avg       0.65      0.64      0.59     34383

------------------------------------------------------------
Confusion Matrix:
[[17877 10549]
 [ 1976  3981]]
```

stchrtr, ruralurbanindicator, lmiindicator, currentapprovalamount, race, businesstype, and payroll_proceed

SMU

04 Results

- Model Performance
- Model Comparisions

Model Performance Accuracy Comparison

	Logistic Regression	CART	Naive Bayes
Overall Accuracy	69%	69%	64%
Non-Community Bank Precision	60%	76%	63%
Community Bank Precision	78%	61%	67%

Model Performance Feature Comparison

	Logistic Regression	CART	Naive Bayes
State Charter	X	X	X
Rural/Urban	X		X
Low-Medium Income	X		X
Approval Amount	X		X
Race	X		X
Business Type	X		X
Payroll Proceed	X		X
Minority		X	
Jobs Reported			

CONCLUSION

Reminder: Our goal was to determine if there exists statistically significant differences or patterns in the **business type**, **business size** (number of jobs), and **business ownership** (e.g., racial minority or gender) for the businesses that received the PPP loans from community banks and/or non-community banks.

✓ **Business Type**

✗ **Business Size**

✓ **Business Ownership**

Next Steps

Other questions to consider...

- Did media perception effect second round loan distribution?
- Are significant variables from the first round the same as the second?
- Would a business that was rejected for a PPP loan through a non-community bank have been able to receive a PPP loan from a community bank?

Potential legislative implications

- Revision of small business standards for federal loan program eligibility
- Special consideration for minority, income, and rural factors
- Expansion of loan education and corporate partnership programs to support small businesses

Thank you! Questions?

Melissa Serrano
serranom@smu.edu

Annalise Sumpon
asumpon@smu.edu

Madi Tedrow
mtedrow@smu.edu

SMU