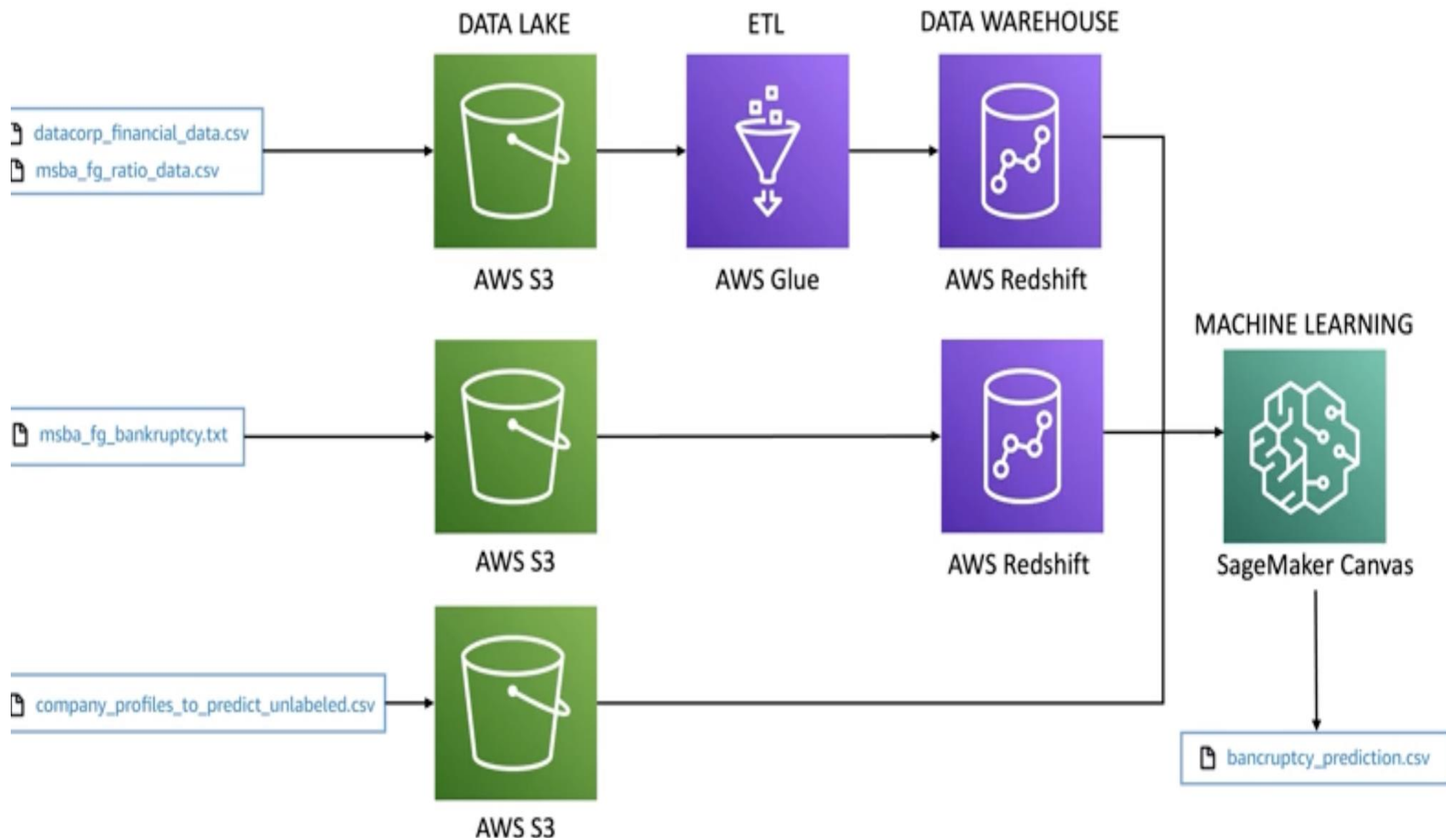




MSBA
FINANCIAL GROUP

DATA ARCHITECHTURE





Amazon S3

- Buckets
- Access Grants
- Access Points
- Object Lambda Access Points
- Multi-Region Access Points
- Batch Operations
- IAM Access Analyzer for S3

Block Public Access settings for this account

Storage Lens

- Dashboards
- Storage Lens groups

Amazon S3 > Buckets

Account snapshot - updated every 24 hours All AWS Regions

Storage lens provides visibility into storage usage and activity trends. [Learn more](#)

View Storage Lens dashboard

General purpose buckets

Directory buckets

General purpose buckets (12) All AWS Regions

Copy ARN

Empty

Delete

Create bucket

Buckets are containers for data stored in S3.

Find buckets by name

	Name	AWS Region	IAM Access Analyzer	Creation date
	msba-fingroup-997	US East (N. Virginia) us-east-1	View analyzer for us-east-1	October 1, 2024, 19:27:08 (UTC-05:00)



[Amazon S3](#) > [Buckets](#) > [msba-fingroup-997](#) > [data_files/](#)

data_files/

Copy S3

Objects

Properties

Objects (3) Info

Copy S3 URI Copy URL Download Open Delete Actions Create folder Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 >

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	datacorp_financial_data.csv	csv	October 1, 2024, 19:31:46 (UTC-05:00)	1.2 MB	Standard
<input type="checkbox"/>	msba_fg_bankruptcy.txt	txt	October 1, 2024, 19:31:45 (UTC-05:00)	53.3 KB	Standard
<input type="checkbox"/>	msba_fg_ratio_data.csv	csv	October 1, 2024, 19:31:45 (UTC-05:00)	614.2 KB	Standard

3

×

la Access Points

Access Points

ons

alyzer for S3

ccess settings for

groups

itions settings

[Amazon S3](#) > [Buckets](#) > [msba-fingroup-997](#) > [prediction/](#)

prediction/

Objects

Properties

Objects (1) Info

Copy S3 URI Copy URL Download Open Delete Actions Create folder

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, grant them permissions. [Learn more](#)

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	company_profiles_to_prediction_unlabeled.csv	csv	October 1, 2024, 19:30:51 (UTC-05:00)	3.3 KB	Standard



AMAZON REDSHIFT



Amazon Redshift



Redshift Serverless [New](#)

Provisioned clusters dashboard

▼ Clusters

Reserved nodes

Snapshots

Query editor

Query editor v2 [↗](#)

Queries and loads

Datashares

Zero-ETL integrations [New](#)

IAM Identity Center connections
[New](#)

► Configurations

📘 Try new Amazon Redshift features in preview.

Create a cluster with preview features. Production use of the cluster is not supported. Use this cluster for testing only.

Create preview cluster



[Amazon Redshift](#) > Clusters

In my account

From other accounts

► Connect to Redshift clusters

Clusters (1) [Info](#)



Query data ▼

Actions ▼

Create cluster

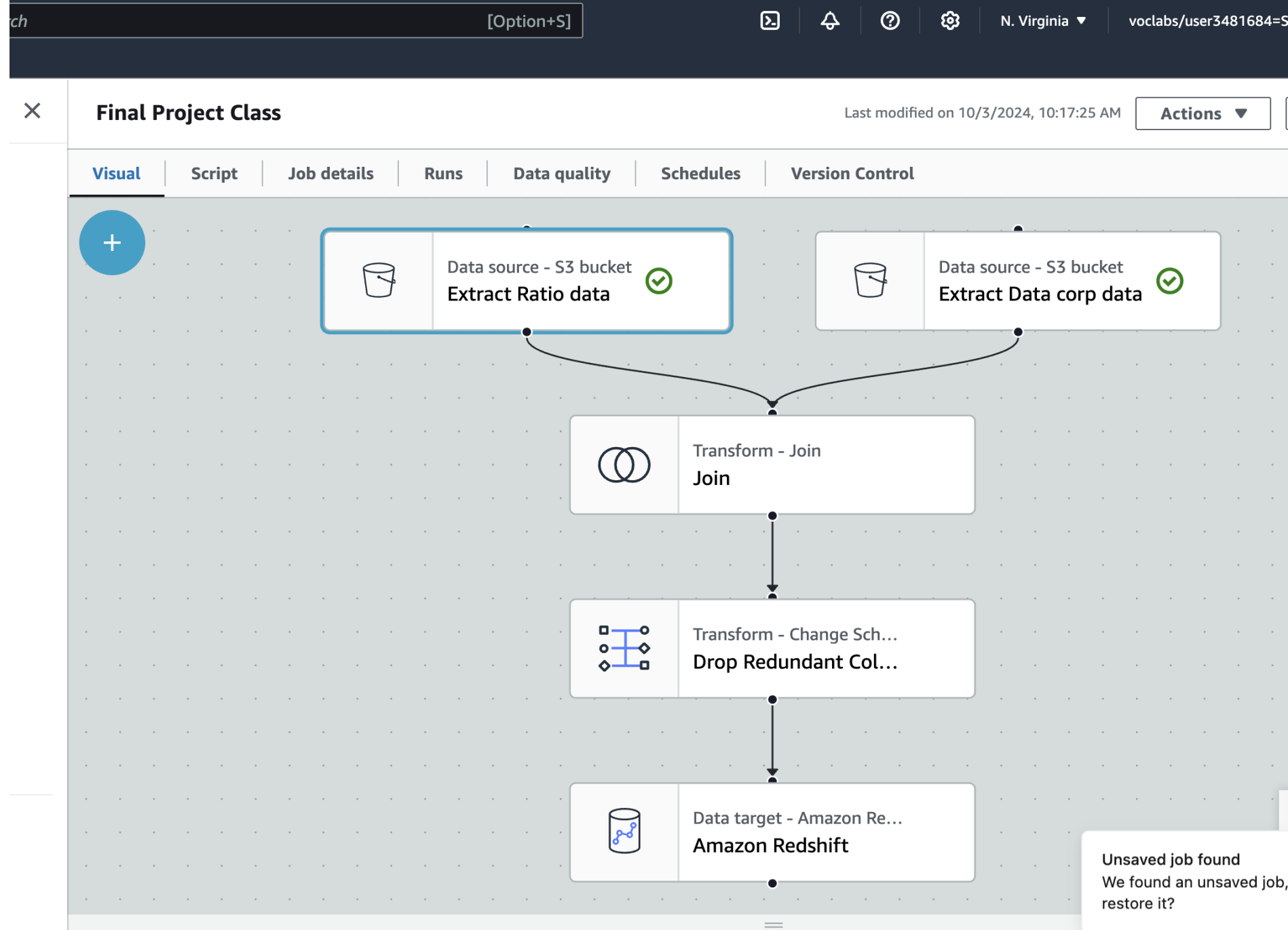
🔍 Find clusters

< 1 > ⚙️

<input type="checkbox"/>	Cluster ▼	Status ▼	Cluster namespace ▼	Availability Zone	Multi-AZ	Storage capacity us..
<input type="checkbox"/>	msba-fingroup-cluster dc2.large 2 nodes 320 GB	✔️ Available	557ddc13-8bc2-4d3c-...	us-east-1a	No	0 %



AWS GLUE FOR ETL





AWS Glue ✕ **Training_set-ETL** Last modified on 10/3/2024, 10:43:29 AM Actions ▾ Save Run

Visual | Script | Job details | Runs | Data quality | Schedules | Version Control

Getting started
ETL jobs
 Visual ETL
 Notebooks
 Job run monitoring
Data Catalog tables
Data connections
Workflows (orchestration)

▼ **Data Catalog**
Databases
 Tables
Stream schema registries
Schemas
Connections
Crawlers
 Classifiers
Catalog settings

► **Data Integration and ETL**

► **Legacy pages**

What's New [↗](#)
Documentation [↗](#)
AWS Marketplace

```
graph TD; A[Data source - Amazon Redshift Extract Bankruptcy] --> C[Transform - Join Join]; B[Data source - Amazon Redshift Extract financials co...] --> C; C --> D[Data target - Amazon Redshift Amazon Redshift];
```




Combining financials_combined and Bankruptcy tables and loading to Redshift Cluster for training data

aws Services s3 All Bookmarks

S3

Amazon Redshift

Redshift Serverless [New](#)

Provisioned clusters dashboard

▼ Clusters

- Reserved nodes
- Snapshots

[Query editor](#)

Query editor v2 [🔗](#)

Queries and loads

Datashares

Zero-ETL integrations [New](#)

IAM Identity Center connections [New](#)

► Configurations

► bankruptcy

► financials_combined_class

▼ training_set

company_id

liability_to_equity

net_income_to_total_assets

working_capital_to_total_assets

net_profit_before_tax_to_paid_in_capital

operating_profit_per_share

net_worth_to_assets

total_expense_to_assets

tax_rate

debt_ratio_percentage

net_income_to_stockholders_equity

retained_earnings_to_total_assets

bankrupt

operating_gross_margin

operating_profit_paid_in_capital

total_asset_turnover

per_share_net_profit_pre_tax

persistent_eps

equity_to_longterm_liability

roa_before_interest_percent_after_tax

working_capital_to_equity

current_liability_to_current_assets

borrowing_dependency

current_liabilities_to_equity

Run

Save

Schedule

Clear

Send feedback

Query results


Table details


No resources


To view details, choose data from navigator.


AMAZON SAGE MAKER


- Importing data from Redshift cluster(training data) and S3 bucket (unlabelled data) into Amazon sagemaker.



Home



Data Wrangler


Datasets


My Models


ML Ops


Ready-to-use


Gen AI

Datasets

There's a new way to join data. You can now join datasets by creating a data flow in Data Wrangler. [Learn how to join data](#)

Import and prepare

Filter by data type:

All

Tabular

Document

Image

Dataset

Import data

Name		Dataset type	Source	Files	Cells (Columns x Rows)	Last updated
<input type="checkbox"/> MSBA-FINAL PROJECT-unlabelled data	V1	Tabular	S3	1	240 (24 x 10)	10/03/2024 8:39
<input type="checkbox"/> MSBA-Final-Project-labelled data	V1	Tabular	Redshift	1	163,656 (24 x 6,819)	10/03/2024 8:41
<input type="checkbox"/> cd_campaign_unlabelled_data	V1	Tabular	S3	1	140 (14 x 10)	09/26/2024 10:22
<input type="checkbox"/> cd_cmapiagn_labeled_data	V1	Tabular	S3	1	617,820 (15 x 41,188)	09/26/2024 10:22
<input type="checkbox"/> canvas-sample-housing.csv	V1	Tabular	S3	1	10,000 (10 x 1,000)	09/23/2024 3:39
<input type="checkbox"/> canvas-sample-diabetic-readmission.csv	V1	Tabular	S3	1	16,000 (16 x 1,000)	09/23/2024 3:39
<input type="checkbox"/> canvas-sample-shipping_logs.csv	V1	Tabular	S3	1	10,000 (10 x 1,000)	09/23/2024 3:40

Building Machine Learning Model

Home

Data Wrangler

Datasets

My Models

ML Ops

Ready-to-use

Gen AI

Help

Log out

My models > MSBA-FINAL-PROJECT-MODEL > Version 1

+ Create new version

Select

Build

Analyze

Predict

Deploy

There's a new way to join data. You can now join datasets by creating a data flow in Data Wrangler. [Learn how to join data](#)

Import and prepare

Select dataset

You can import a tabular dataset or choose one that has already been imported. Your dataset must contain at least one input column and a target column.

Name	Columns	Rows	Cells	Created	Status
<div><div></div><div>MSBA-Final-Project-labelled data</div><div>V1</div></div>	24	6,819	163,656	10/03/2024 8:03 PM	Ready

My models > MSBA-FINAL-PROJECT-MODEL > Version 1

+ Create new version

Select

Build

Analyze

Predict

Deploy

Select a column to predict

Choose the target column. The model that you build predicts values for the column that you select.

Target column

bankrupt

Value distribution

0

1

Model type

SageMaker Canvas automatically recommends the appropriate model type for your analysis.

2 category prediction

Your model classifies bankrupt into two categories.

Configure model

Quick build

Preview model

MSBA-Final-Project-labelled data

Full dataset: 6.8k rows

Data visualizer

Column name ↓	Data type ⓘ	Feature type ⓘ	Missing ⓘ	Mismatched ⓘ	Unique ⓘ	Mode ⓘ
working_capital_to_total_assets	Numeric	-	0.00% (0)	0.00% (0)	6,819	0
working_capital_to_equity	Numeric	-	0.00% (0)	0.00% (0)	6,818	0.74
total_expense_to_assets	Numeric	-	0.00% (0)	0.00% (0)	6,819	0
total_asset_turnover	Numeric	-	0.00% (0)	0.00% (0)	381	0.08
tax_rate	Numeric	-	0.00% (0)	0.00% (0)	2,488	0

Total columns: 24

Total rows: 6,819

Total cells: 163,656

Show dropped columns

MODEL PREDICTIONS

Home

Data Wrangler

Datasets

My Models

ML Ops

Ready-to-use

Gen AI

Help

Log out

My models > MSBA-FINAL-PROJECT-MODEL > Version 1

+ Create new version

Select

Build

Analyze

Predict

Deploy

Model status

Quick build

Accuracy

F1

Optimization metric

95.161%

0.468

Predict

Standard build

Deploy

The model predicts the correct Bankrupt 95.161% of the time.

Overview

Scoring

Advanced metrics

Model leaderboard

Positive Class

F1

Optimization metric

Accuracy

Precision

Recall

AUC-ROC

1

0

46.774%

95.161%

36.25%

65.909%

0.927

Metrics table

Confusion matrix

Precision recall curve

Metrics table

Metric name

Value

precision

0.362

recall

0.659

accuracy

0.952

MSBA-Final-Project-labelled data

Total columns: 24

Total rows: 6,819

Total cells: 163,656

bankrupt

2 category prediction

Predict

MODEL PREDICTIONS

Home

Data Wrangler

Datasets

My Models

ML Ops

Ready-to-use

Gen AI

Help

Log out

My models > MSBA-FINAL-PROJECT-MODEL > Version 1

+ Create new version

↺

⋮

Select

Build

Analyze

Predict

Deploy

Model status

Quick build

Accuracy ⓘ

F1 ⓘ Optimization metric

95.161%

0.468

The model predicts the correct Bankrupt 95.161% of the time. ⓘ

Predict

Standard build

Deploy

Overview

Scoring

Advanced metrics

Model leaderboard

All predictions

Predicted

Actual

Total 1364

0

Correct 0 1320

Incorrect

Model accuracy insights

If the model predicts 0, it is correct 98.832% of the time. ⓘ

For the values that are 0 in the dataset, the model predicted 96.136% of them to be 0. ⓘ

MSBA-Final-Project-labelled data

Total columns: 24

Total rows: 6,819

Total cells: 163,656

bankrupt

2 category prediction

Predict

My models > MSBA-FINAL-PROJECT-MODEL > Version 1

+ Create new version

↺

⋮

Select

Build

Analyze

Predict

Deploy

Model status

Quick build

Accuracy ⓘ

F1 ⓘ Optimization metric

95.161%

0.468

The model predicts the correct Bankrupt 95.161% of the time. ⓘ

Predict

Standard build

Deploy

Overview

Scoring

Advanced metrics

Model leaderboard

Positive Class

F1 ⓘ Optimization metric

Accuracy ⓘ

Precision ⓘ

Recall ⓘ

AUC-ROC ⓘ

1 0 46.774%

95.161%

36.25%

65.909%

0.927

Metrics table

Confusion matrix

Precision recall curve

Actual values

1

0

True positive (TP)

29

2.1% of predicted results

False positive (FP)

51

3.7% of predicted results

False negative (FN)

15

1.1% of predicted results

True negative (TN)

1269

93.0% of predicted results

Exploratory Data Analysis(Finding 1)

models > MSBA-FINAL-PROJECT-MODEL > Version 1

+ Create new version

Select Build **Analyze** Predict Deploy

Model status [Quick build](#)

Accuracy ⓘ F1 ⓘ Optimization metric

95.161% 0.468

The model predicts the correct Bankrupt 95.161% of the time. ⓘ

Predict

Standard build

Deploy

Overview Scoring Advanced metrics

Model leaderboard

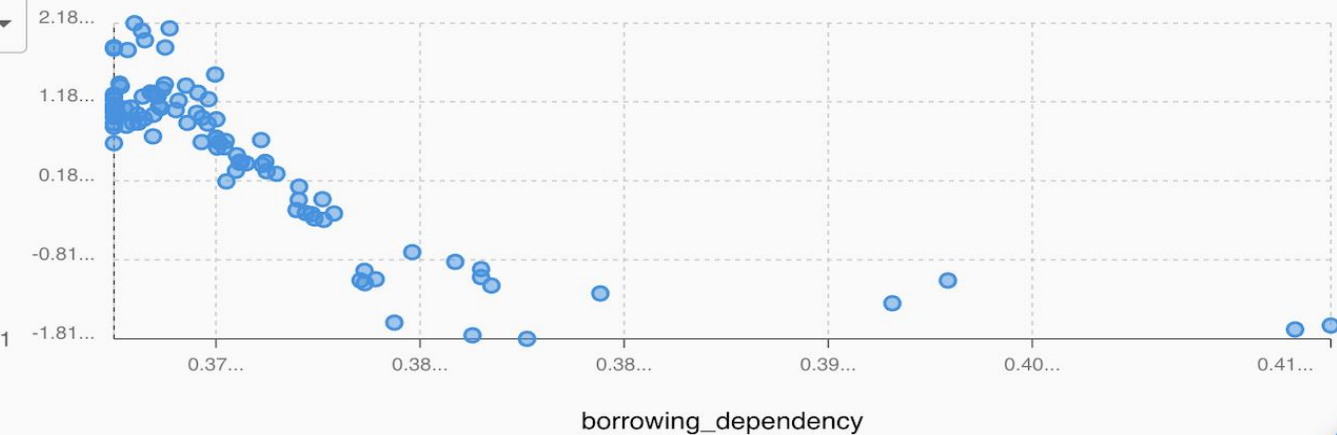
Column impact ⓘ ↓

Search columns...

1	<u>borrowing_dependency</u>	16.211%
2	<u>persistent_eps</u>	12.402%
3	<u>retained_earnings_to_total_assets</u>	11.501%
4	<u>net_income_to_total_assets</u>	10.036%
5	<u>current_liability_to_current_assets</u>	6.873%

Impact on prediction

Impact of borrowing_dependency on prediction of bankrupt



EXPLORATORY DATA ANALYSIS (FINDING 2)

Overview

Scoring

Advanced metrics

Model leaderboard



Column impact ⓘ ↓

Search columns...

1 borrowing_dependency 16.211%

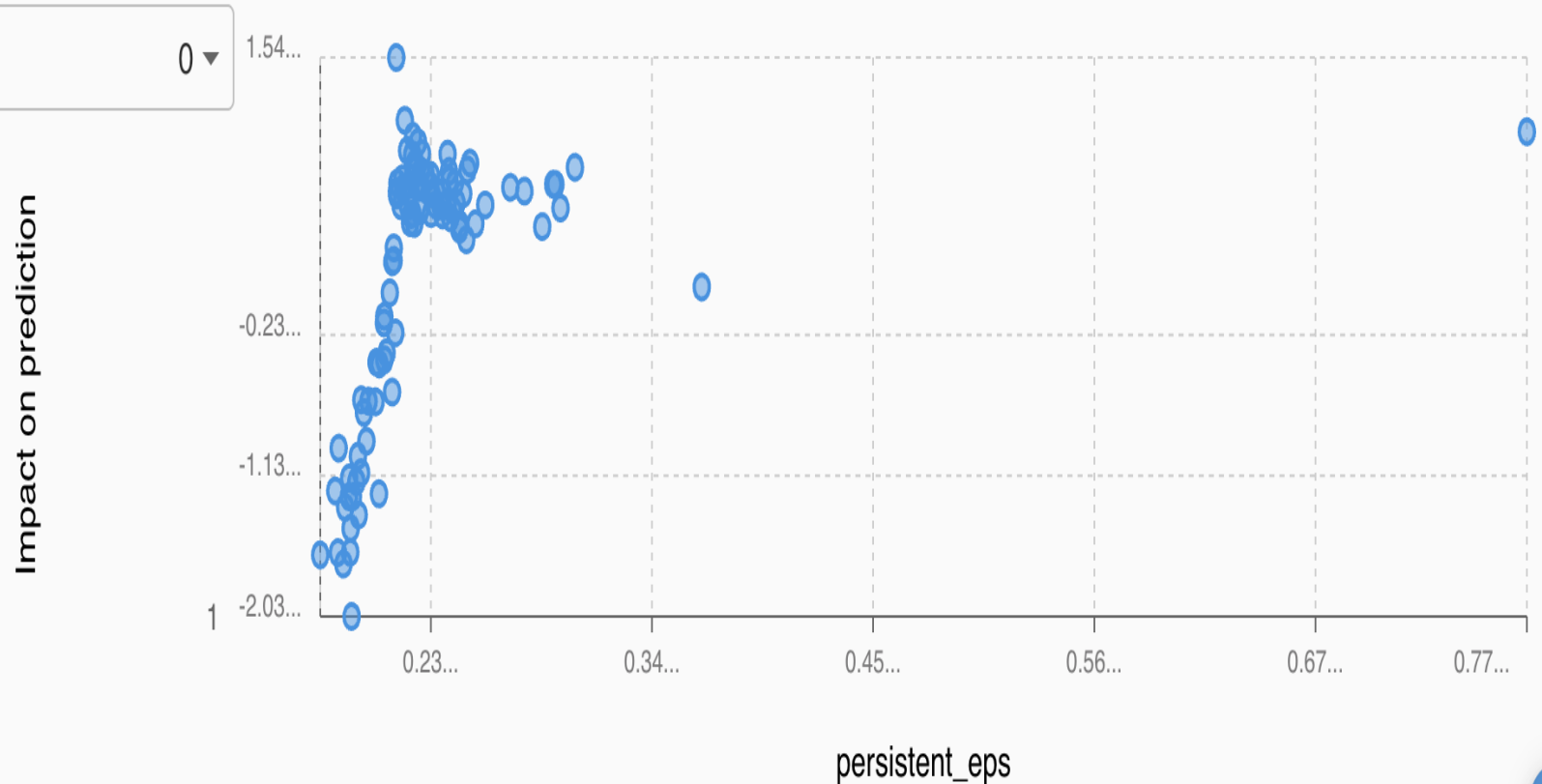
2 persistent_eps 12.402%

3 retained_earnings_to_total_assets 11.501%

4 net_income_to_total_assets 10.036%

5 current_liability_to_current_assets 6.872%

Impact of persistent_eps on prediction of bankrupt



Prediction (bankrupt)	Probability	borrowing_d...	company	company_id	current_liabi...	current_liabi...	debt_ratio_p...
1	99.2%	0.458818609	western corp	id_6988	0.372217526	0.060766259	0.269038909
1	97.4%	0.37930429	design solutions	id_7413	0.333345174	0.041219716	0.16186474
0	79.4%	0.384998982	innocore	id_8801	0.337392013	0.060765125	0.216101823
0	99.9%	0.374219105	pharmasolve	id_9614	0.329803726	0.030201105	0.108202074
0	99.9%	0.370253398	ninetech	id_9131	0.328899756	0.021711111	0.053590561
0	99.7%	0.37450876	songster inc	id_7102	0.330409488	0.025494302	0.121292741
0	99.8%	0.374179962	rogers and sons	id_7612	0.327484654	0.047166548	0.103576503
0	98.8%	0.373113046	Hallandall ag.	id_9904	0.33032211	0.03711102	0.02185827
0	99.5%	0.377306898	Foster & Kruse	id_6905	0.331114215	0.04351399	0.144662454
0	99.5%	0.37315107	Highwood & Hart	id_8039	0.329035251	0.028901672	0.10656952

Predictions of our
model for the un-
labelled data

TOP 5 PREDICTORS FOR COMPANY GOING BANKRUPT

bankrupt	probability	company	company_id	borrowing_dependency	persistent_eps	Retained_earnings_to_total_assets	net_income_to_total_assets	current_liability_to_current_assets
1	0.99288565	western corp	id_6988	0.458818609	0.15429706	0.8794444615	0.64180449	0.060766259
1	0.97423822	design solutions	id_7413	0.37930429	0.217736598	0.933467259	0.800780431	0.041219716
0	0.7941519	innocore	id_8801	0.384998982	0.161482462	0.911441408	0.736619098	0.060765125
0	0.99912834	pharmasolve	id_9614	0.374219105	0.225205635	0.935449309	0.815350392	0.030201105
0	0.9995966	ninetech	id_9131	0.370253398	0.218398412	0.935200209	0.80364729	0.021710461
0	0.99780607	songster inc	id_7102	0.37450876	0.217831143	0.938289897	0.804194657	0.025494302
0	0.99848729	rogers and sons	id_7012	0.374179962	0.21896568	0.942828956	0.814110614	0.047166348
0	0.98857784	Hallandall ag.	id_9904	0.373113046	0.20468942	0.932955215	0.804886953	0.033711602
0	0.99519491	Foster & Kruse	id_6905	0.377306898	0.219343859	0.930139635	0.803260161	0.04351399
0	0.9956789	Highwood & Hart	id_8039	0.37315107	0.212914815	0.930763203	0.794157853	0.028901672



RECOMMENDATIONS BASED ON THE PREDICTIONS

- I would suggest investing in the companies Pharmasolve , Nine-Tech, and Rodgers and Sons as all three show a very high probability of the companies not being bankrupt.
- Additionally, as per the top 5 attributes specified in the previous slide suggests that these companies have a lower borrowing dependency and high persistent earning per share when compared to the other companies.

CONCLUSIONS

- As per the model predictions, the following companies are at the highest probability of going bankrupt.
 - -> Western Corp
 - -> Design Solutions
- The companies Pharmasolve , Nine-Tech, and Rodgers and Sons are safe to invest in, as the model predicts a high chance for not going bankrupt.

