

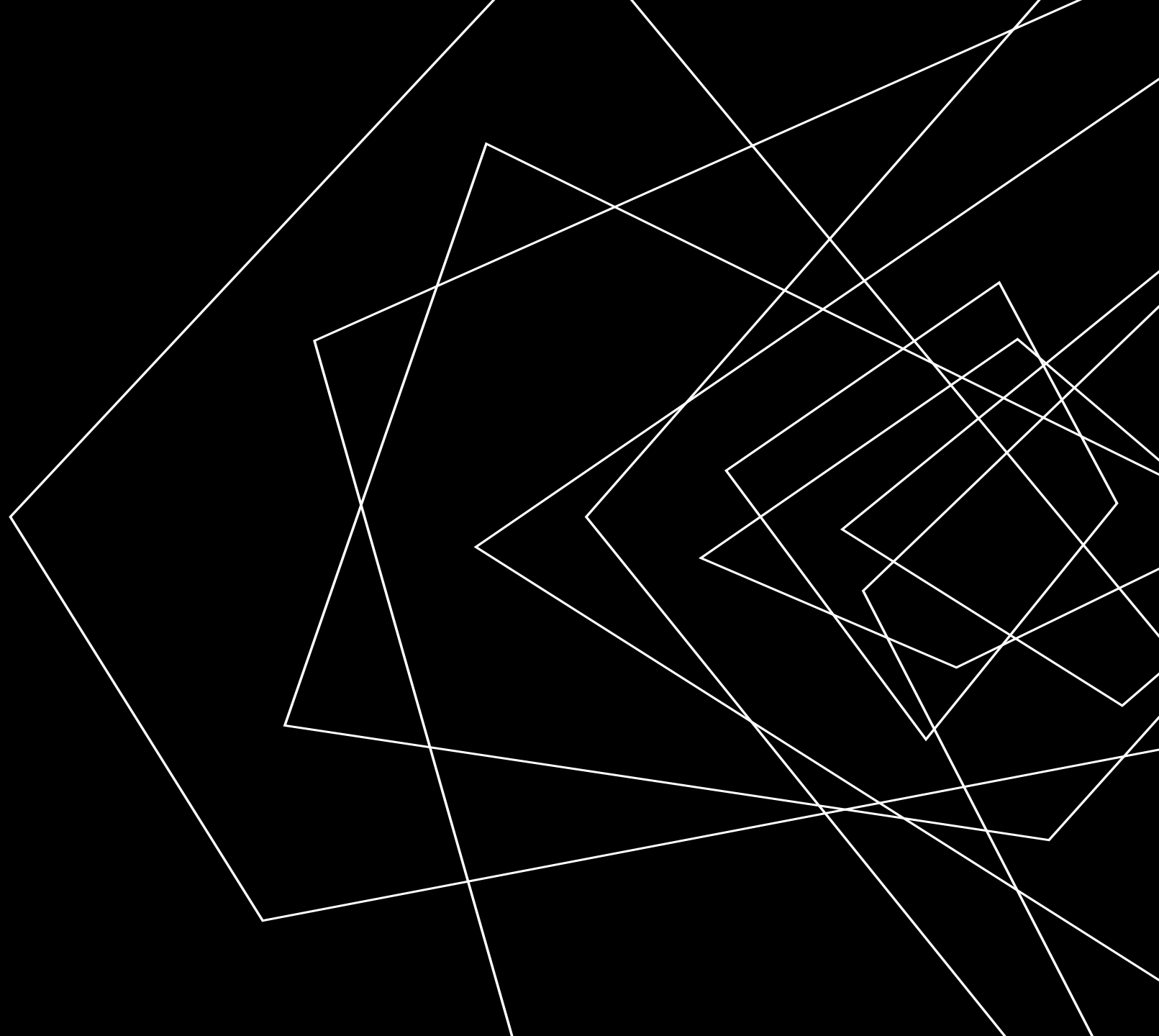
Abstract geometric lines in black on a white background, forming various overlapping polygons and shapes.

MSBA FINANCIAL GROUP PROJECT

AMBIKA BURRAMUKKU

AGENDA

- Overview of the Data Flow Diagram
- Examine each step
- Explore findings
- Review companies



Data Architecture



msba_fg_bankruptcy.txt
msba_fg_ratios.csv
datacorp_financials.csv

Microsoft Excel

Data Lake



bucket: msba-fingroup-445

AWS S3

ETL



AWS Glue

Glue job: Financial_ETL

Data Warehouse



AWS Redshift

Cluster: msba-fingroup-cluster
Tables: bankruptcy,
financials_combined,
training_set_2

Machine Learning



AWS Sagemaker

bankruptcy_prediction.csv

Model: MSBA-finalgroup-model
company_profiles_to_
predict_unlabeled.csv

Data Architecture



msba_fg_bankruptcy.txt
msba_fg_ratios.csv
datacorp_financials.csv

Microsoft Excel

bankruptcy_prediction.csv

Machine Learning



Cluster: msba-fingroup-cluster
Tables: bankruptcy,
financials_combined,
training_set_2

AWS Sagemaker

Data Lake



AWS S3

bucket: msba-fingroup-445

Data Warehouse



AWS Redshift

ETL



AWS Glue

Glue job: Financial_ETL

Model: MSBA-finalgroup-model
company_profiles_to_
predict_unlabeled.csv

Data Architecture



msba_fg_bankruptcy.txt
msba_fg_ratios.csv
datacorp_financials.csv

Microsoft Excel



Data Lake



AWS S3

bucket: msba-fingroup-445



ETL



AWS Glue

Glue job: Financial_ETL



Data Warehouse



AWS Redshift



Machine Learning



Cluster: msba-fingroup-cluster
Tables: bankruptcy,
financials_combined,
training_set_2

AWS Sagemaker

bankruptcy_prediction.csv



Model: MSBA-finalgroup-model
company_profiles_to_
predict_unlabeled.csv

Objects (2) 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 objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

| <input type="checkbox"/> | Name | Type | Last modified |
|--------------------------|-------------|--------|---------------|
| <input type="checkbox"/> | data_files/ | Folder | - |
| <input type="checkbox"/> | prediction/ | Folder | - |

Objects (1) Info

Copy S3 URI

Copy URL

Download

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) objects, you'll need to explicitly 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_predict_unlabeled.csv | csv | October 2, 2024, 22:50:59 (UTC-05:00) | 3.3 KB | Standard |

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

| <input type="checkbox"/> | Name | Type | Last modified | Size | Storage class |
|--------------------------|-----------------------------|------|---------------------------------------|----------|---------------|
| <input type="checkbox"/> | datacorp_financial_data.csv | csv | October 2, 2024, 22:46:18 (UTC-05:00) | 1.2 MB | Standard |
| <input type="checkbox"/> | msba_fg_bankruptcy.txt | txt | October 2, 2024, 22:46:19 (UTC-05:00) | 53.3 KB | Standard |
| <input type="checkbox"/> | msba_fg_ratio_data.csv | csv | October 2, 2024, 22:46:20 (UTC-05:00) | 614.2 KB | Standard |

Data Architecture



msba_fg_bankruptcy.txt
msba_fg_ratios.csv
datacorp_financials.csv

Microsoft Excel

bankruptcy_prediction.csv

Machine Learning



Cluster: msba-fingroup-cluster
Tables: bankruptcy,
financials_combined,
training_set_2

AWS Sagemaker

Data Lake



AWS S3

bucket: msba-fingroup-445

Data Warehouse



AWS Redshift

ETL



AWS Glue
Glue job: Financial_ETL

Visual

Script

Job details

Runs

Data quality

Schedules

Version Control



Data source - S3 bucket

Ratio Data

Data source - S3 bucket

DATA CORPS

Transform - Join

Join

Transform - Change Sch...

Change Schema

Data target - Amazon Re...

Amazon Redshift



Data Architecture



msba_fg_bankruptcy.txt
msba_fg_ratios.csv
datacorp_financials.csv

Microsoft Excel

Data Lake



bucket: msba-fingroup-445

AWS S3

ETL



AWS Glue

Glue job: Financial_ETL

Data Warehouse



AWS Redshift

Machine Learning



Cluster: msba-fingroup-cluster
Tables: bankruptcy,
financials_combined,
training_set_2

AWS Sagemaker

bankruptcy_prediction.csv

Model: MSBA-finalgroup-model
company_profiles_to_
predict_unlabeled.csv

Resources [Info](#)

Select database [Info](#)

To view schemas, select a database.

dev

Select schema [Info](#)

To view tables, select a schema.

public

Filter tables

1

► bankruptcy

...

► financials_combined

...

► training_set

...

► training_set_2

...

Status [Connected](#)

database dev user msbauser

Change connection

Query 2

Query 3

Query 1

Query 4

Query 5

+

▼

↶ ↷ @ ⌵ /* 📄 🔍

1 CREATE TABLE financials_combined (
2 company_id VARCHAR(10)
3);
4
5 CREATE TABLE bankruptcy
6 (
7 Company_ID VARCHAR(10),
8 Bankrupt INTEGER
9);
10
11 create table training_set
12 (
13 company_id VARCHAR(10)
14);

Run

Save

Schedule

Clear

Send feedback

Resources [Info](#)

Select database [Info](#)

To view schemas, select a database.

dev

Select schema [Info](#)

To view tables, select a schema.

public

Filter tables

1

► bankruptcy

...

Status [Connected](#)

database dev user msbauser

Change connection

Query 2

Query 3

Query 1

Query 4

Query 5

+

▼

↶ ↷ @ ⌵ /* 📄 🔍

1 COPY bankruptcy
2 FROM 's3://msba-fingroup-445/data_files/msba_fg_bankruptcy.txt'
3 CREDENTIALS 'aws_iam_role=arn:aws:iam::606620512111:role/LabRole'
4 DELIMITER '|'
5 IGNOREHEADER 1
6 ;
7
8 UPDATE financials_combined
9 SET "company_id" = CONCAT('id_', "company_id");
10
11

```

1 create table training_set_2 as
2 select
3 b.company_id,
4 b.bankrupt,
5 f.liability_to_equity,
6 f.net_income_to_total_assets,
7 f.working_capital_to_total_assets,
8 f.net_profit_before_tax_to_paid_in_capital,
9 f.operating_profit_per_share,
10 f.net_worth_to_assets,
11 f.total_expense_to_assets,
12 f.tax_rate,
13 f.debt_ratio_percentage,
14 f.net_income_to_stockholders_equity,
15 f.retained_earnings_to_total_assets,
16 f.operating_gross_margin,
17 f.operating_profit_paid_in_capital,
18 f.total_asset_turnover,
19 f.per_share_net_profit_pre_tax,
20 f.persistent_eps,
21 f.equity_to_longterm_liability,
22 f.roa_before_interest_percent_after_tax,
23 f.current_liability_to_current_assets,
24 f.working_capital_to_equity,
25 f.borrowing_dependency,
26 f.current_liabilities_to_equity
27 from bankruptcy b
28 left join financials_combined f
29 on b.company_id = f.company_id ;
30
31 select * from training_set_2 Limit 10

```

Query results

Table details

Query 1009434 [🔗](#)

Execution

Data

Visualize

✓ Completed, started on October 06, 2024 at 13:29:17
ELAPSED TIME: 00 m 02 s

Rows returned (10)

Export ▼

🔍 Search rows

< 1 > ⚙️

| company_id ▼ | bankrupt ▼ | liability_to_equity ▼ | net_income_to_total_assets ▼ | working_capital_to_total ▼ |
|--------------|------------|-----------------------|------------------------------|----------------------------|
| id_3620 | 0 | 0.280122394 | 0.815155163 | 0.794416249 |
| id_3620 | 0 | 0.280122394 | 0.815155163 | 0.794416249 |
| id_4904 | 0 | 0.279871658 | 0.801365089 | 0.786462568 |

Data Architecture



msba_fg_bankruptcy.txt
msba_fg_ratios.csv
datacorp_financials.csv

Microsoft Excel

bankruptcy_prediction.csv

Machine Learning



AWS Sagemaker

Cluster: msba-fingroup-cluster
Tables: bankruptcy,
financials_combined,
training_set_2

Data Lake



AWS S3

bucket: msba-fingroup-445

Data Warehouse



AWS Redshift

ETL



AWS Glue

Glue job: Financial_ETL

Model: MSBA-finalgroup-model
company_profiles_to_
predict_unlabeled.csv



Home



Data Wrangler



Datasets



My Models



ML Ops



Ready-to-use



Gen AI



Help



Log out

My models > MSBA-fingroup-model > Version 1

+ Create new version



Select

Build

Analyze

Predict

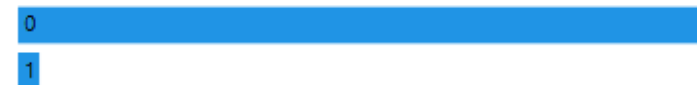
Deploy

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



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-fingroup-labeled-data

Random sample: 20.0k rows

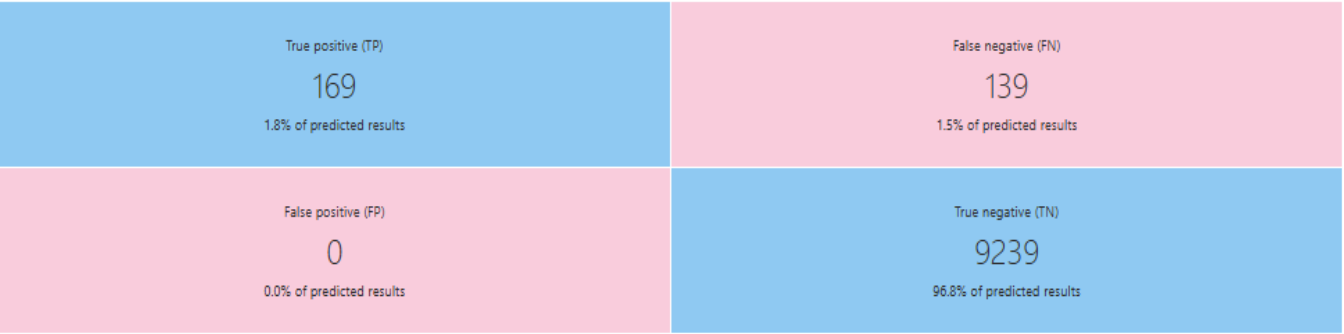


[Data visualizer](#)



| working_capital_to_total_... 123 | working_capital_to_equity 123 | total_expense_to_assets 123 | total_asset_turnover 123 | tax_rate 123 | roa_before_interest_perc... 123 | retained_earnings_to_tot... 123 | per_share_net_profit_pre_... |
|----------------------------------|-------------------------------|-----------------------------|--------------------------|--------------|---------------------------------|---------------------------------|------------------------------|
| | | | | | | | |
| 0.48 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | 0.00 | 0.01 |
| 0.98 | 0.75 | 0.11 | 0.52 | 0.63 | 0.80 | 0.99 | |
| 0.786462568 | 0.734540657 | 0.020215473 | 0.167916042 | 0.260350238 | 0.546282163 | 0.937504008 | 0.174738269 |
| 0.786462568 | 0.734540657 | 0.020215473 | 0.167916042 | 0.260350238 | 0.546282163 | 0.937504008 | 0.174738269 |
| 0.765292062 | 0.733147267 | 0.012396586 | 0.092953523 | 0.182508351 | 0.549552987 | 0.93900422 | 0.183851774 |
| 0.662359963 | 0.7145055 | 0.020195192 | 0.137931034 | 0 | 0.538813781 | 0.91282594 | 0.167959629 |

Total columns: 24 Total rows: 47,733 Total cells: 1,145,592 Previewing first 100 rows ☒ Show dropped columns



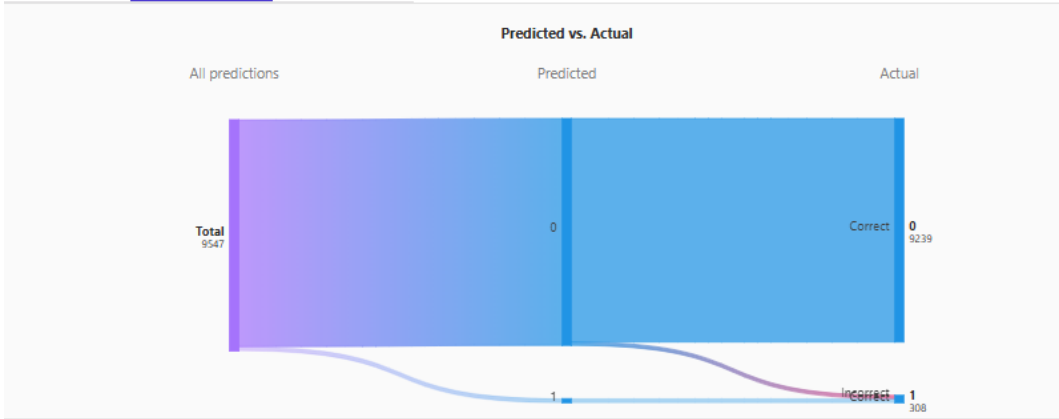
Model status [Quick build](#)

Accuracy ⓘ F1 ⓘ Optimization metric

98.544% 0.709

The model predicts the correct Bankrupt 98.544% of the time. ⓘ

Overview Scoring Advanced metrics



Model status [Quick build](#)

Accuracy ⓘ F1 ⓘ Optimization metric

98.544% 0.709

The model predicts the correct Bankrupt 98.544% of the time. ⓘ

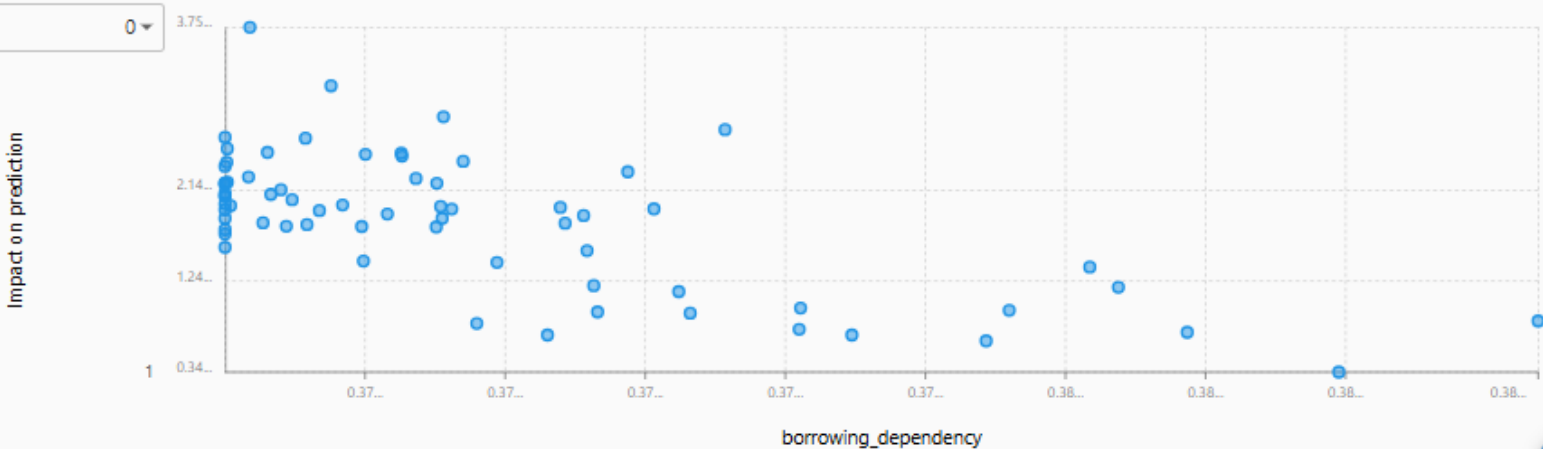
Overview Scoring Advanced metrics

Column impact ⓘ ↓

Search columns...

- 1 borrowing_dependency 16.84%
- 2 retained_earnings_to_total_assets 11.351%
- 3 net_income_to_total_assets 8.462%
- 4 current_liabilities_to_equity 7.075%
- 5 current_liability_to_current_assets 5.343%
- 6 working_capital_to_equity 4.574%

Impact of borrowing_dependency on prediction of bankrupt



Data Architecture



msba_fg_bankruptcy.txt
msba_fg_ratios.csv
datacorp_financials.csv

Microsoft Excel

bankruptcy_prediction.csv

Machine Learning



Cluster: msba-fingroup-cluster
Tables: bankruptcy,
financials_combined,
training_set_2

AWS Sagemaker

Data Lake



AWS S3

bucket: msba-fingroup-445

Data Warehouse



AWS Redshift

ETL



AWS Glue

Glue job: Financial_ETL

Model: MSBA-finalgroup-model
company_profiles_to_
predict_unlabeled.csv

| bankrupt | probability | borrowing | company | company_ | current_li | current_li | debt_ratio | equity_to_l | liability_to | net_incom |
|----------|-------------|-----------|------------------|----------|------------|------------|------------|-------------|--------------|-----------|
| 1 | 99.7843% | 0.45882 | western corp | id_6988 | 0.37222 | 0.06077 | 0.26904 | 0.21688 | 0.33732 | 0.76597 |
| 1 | 99.6527% | 0.3793 | design solutions | id_7413 | 0.33335 | 0.04122 | 0.16186 | 0.12081 | 0.28276 | 0.84053 |
| 0 | 99.5934% | 0.385 | innocore | id_8801 | 0.33739 | 0.06077 | 0.2161 | 0.12056 | 0.2925 | 0.82998 |
| 0 | 99.9958% | 0.37422 | pharmasolve | id_9614 | 0.3298 | 0.0302 | 0.1082 | 0.11451 | 0.27861 | 0.84146 |
| 0 | 99.9987% | 0.37025 | ninetech | id_9131 | 0.32809 | 0.02171 | 0.05859 | 0.11093 | 0.27642 | 0.84049 |
| 0 | 99.9933% | 0.37451 | songster inc | id_7102 | 0.33041 | 0.02549 | 0.12129 | 0.1155 | 0.27939 | 0.84069 |
| 0 | 99.9944% | 0.37418 | rogers and sons | id_7012 | 0.32748 | 0.04717 | 0.10358 | 0.12094 | 0.27836 | 0.84134 |
| 0 | 99.9455% | 0.37311 | Hallandall ag. | id_9904 | 0.32804 | 0.03371 | 0.09439 | 0.11646 | 0.27789 | 0.84065 |
| 0 | 99.9503% | 0.37731 | Foster & Kruse | id_6905 | 0.33111 | 0.04351 | 0.14466 | 0.11696 | 0.28111 | 0.8407 |
| 0 | 99.9851% | 0.37315 | Highwood & Hart | id_8039 | 0.32904 | 0.0289 | 0.10657 | 0.11357 | 0.27852 | 0.83991 |

- Companies to consider:

- Innocore
- Pharmasolve
- Ninetech
- Songster Inc
- Rogers and Sons
- Hallandall ag.
- Foster & Kruse
- Highwood & Hart

A series of white, thin, overlapping geometric lines on a black background, forming a complex, abstract shape on the left side of the slide.

THANK YOU