US Company Bankruptcy Detection

AASD4010 Deep Learning I

Group 1

Chotiros Srisiam	#101411914
Kajhonprom Trongkitroongruang	#101446812
Pat Boonprasertsri	#101410612
Pek Chansatit	#101439953
Vitchaya Siripoppohn	#101481464

Table of Contents

01 Team Members

Roles


Kajhonprom T.
Subject Matter Expert


Chotiros S.
Project Manager
Data Analyst


Vitchaya S.
Model Architect


Pat B.
Data Analyst
QA Analyst


Pek C.
Data Scientist

Project Overview

02

Motivation

Look at the real case of the American energy company that importance of a deep learning model for bankruptcy prediction

Enron Corporation

Enron Corporation, once a global energy giant, collapsed in 2001 due to widespread accounting fraud and financial mismanagement.

Original Value (2001)
= $70 billion
Adjusted Value (2024)
= $107 billion

Value Or Cost Of The Damage Incurred In The Enron Scandal

1. Market Capitalization Loss
2. Investor Losses
3. Employee Impact
4. Legal and Regulatory Consequences
5. Collapse of Arthur Andersen
6. Broader Economic Impact
7. Total Financial Fallout

Traditional Analysis

Traditional financial analysis methods **failed to detect Enron's fraudulent** activities, as the company manipulated financial statements and hid debt through off-balance-sheet entities.

Deep Learning Model

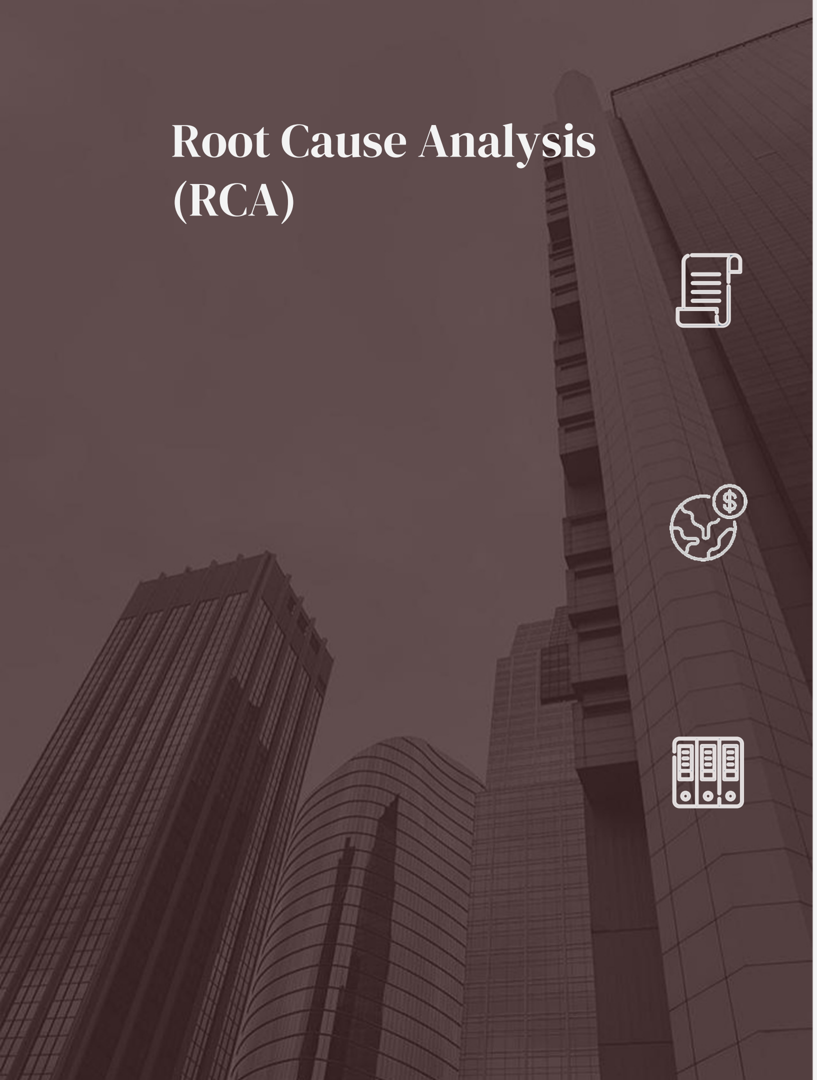
A deep learning model for bankruptcy prediction could have played a crucial role in identifying Enron's financial distress early on by:

1. **Analysis** a broader range of financial indicators and patterns.

2. **Recognizing** anomalies and irregularities in real-time data.

3. **Providing** a proactive mechanism for stakeholders to address risks promptly.

Root Cause Analysis (RCA)



Factors contributing to the need for a bankruptcy prediction model

Financial Statement Complexity

Complex financial statements pose challenges for manual analysis. Deep learning models uncover hidden patterns in intricate financial data, providing early warning signs beyond human capabilities.

Global Economic Volatility

Economic uncertainties and market fluctuations heighten business vulnerability. A predictive model, considering macroeconomic indicators, offers a comprehensive view, aiding companies in preparing for economic downturns.

Data Overload and Processing Speed

Swift processing of vast financial data is crucial. Deep learning models, adept at handling large datasets, emerge as essential tools for timely decision-making.

Our Project

Predicting bankruptcy in American public companies is vital for decision-makers. This project develops a precise model to assist investors in risk management.

03 Dataset

About Dataset

company_na	status_label	year	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18
C_1	alive	1999	511.267	833.107	18.373	89.031	336.018	35.163	128.348	372.7519	1024.333	740.998	180.447	70.658	191.226	163.816	201.026	1024.333	401.483	935.302
C_1	alive	2000	485.856	713.811	18.577	64.367	320.59	18.531	115.187	377.118	874.255	701.854	179.987	45.79	160.444	125.392	204.065	874.255	361.642	809.888
C_1	alive	2001	436.656	526.477	22.496	27.207	286.588	-58.939	77.528	364.5928	638.721	710.199	217.699	4.711	112.244	150.464	139.603	638.721	399.964	611.514
C_1	alive	2002	396.412	496.747	27.172	30.745	259.954	-12.41	66.322	143.3295	606.337	686.621	164.658	3.573	109.59	203.575	124.106	606.337	391.633	575.592
C_1	alive	2003	432.204	523.302	26.68	47.491	247.245	3.504	104.661	308.9071	651.958	709.292	248.666	20.811	128.656	131.261	131.884	651.958	407.608	604.467
C_1	alive	2004	474.542	598.172	27.95	61.774	255.477	15.453	127.121	522.6794	747.848	732.23	227.159	33.824	149.676	160.025	142.45	747.848	417.486	686.074
C_1	alive	2005	624.454	704.081	29.222	91.877	323.592	35.163	136.272	882.6283	897.284	978.819	318.576	62.655	193.203	187.788	183.55	897.284	556.102	805.407
C_1	alive	2006	645.721	837.171	32.199	118.907	342.593	58.66	181.691	1226.1925	1061.169	1067.633	253.611	86.708	223.998	256.506	242.153	1061.169	573.39	942.262

78,682 instances

8,971 companies

Time series

1999-2018

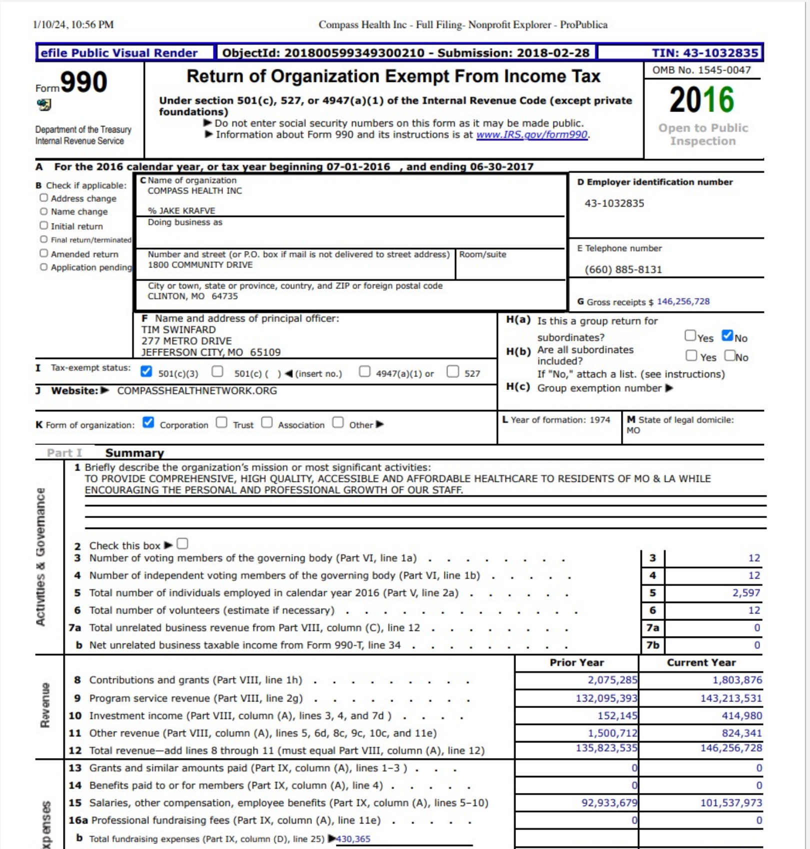
21 attributes

- Company
- Company Status
- Year
- Current assets
- Cost of goods sold
- Depreciation and amortization
- EBITDA - Earnings before interest, taxes, depreciation, and amortization.
- Inventory
- Net Income
- Total Receivables
- Market value
- Net sales
- Total assets
- Total Long-term debt
- EBIT - Earnings before interest and taxes.
- Gross Profit
- Total Current Liabilities
- Retained Earnings
- Total Revenue
- Total Liabilities
- Total Operating Expenses

Reference: Accounting data of NYSE and NASDAQ companies (1999-2018), Collected by Utkarsh Singh.
(https://www.kaggle.com/datasets/utkarshx27/american-companies-bankruptcy-prediction-dataset)

About Data

The dataset labels a company as bankrupt (1) if it files for Chapter 11 (reorganization) or Chapter 7 (complete cessation). The fiscal year before such filing is labeled "Bankruptcy." Otherwise, it is considered operating normally (0) according to SEC definitions.



1/10/24, 10:56 PM Compass Health Inc - Full Filing- Nonprofit Explorer - ProPublica

efile Public Visual Render | ObjectId: 201800599349300210 - Submission: 2018-02-28 | TIN: 43-1032835

Form 990

Department of the Treasury
Internal Revenue Service

Return of Organization Exempt From Income Tax

Under section 501(c), 527, or 4947(a)(1) of the Internal Revenue Code (except private foundations)

▶ Do not enter social security numbers on this form as it may be made public.
▶ Information about Form 990 and its instructions is at www.IRS.gov/form990.

OMB No. 1545-0047

2016

Open to Public Inspection

A For the 2016 calendar year, or tax year beginning 07-01-2016 , and ending 06-30-2017

B Check if applicable:
☐ Address change
☐ Name change
☐ Initial return
☐ Final return/terminated
☐ Amended return
☐ Application pending

C Name of organization
COMPASS HEALTH INC

% JAKE KRAFVE
Doing business as

Number and street (or P.O. box if mail is not delivered to street address) Room/suite
1800 COMMUNITY DRIVE

City or town, state or province, country, and ZIP or foreign postal code
CLINTON, MO 64735

D Employer identification number
43-1032835

E Telephone number
(660) 885-8131

G Gross receipts $ 146,256,728

F Name and address of principal officer:
TIM SWINFARD
277 METRO DRIVE
JEFFERSON CITY, MO 65109

H(a) Is this a group return for subordinates? ☐ Yes ☑ No
H(b) Are all subordinates included? ☐ Yes ☐ No
If "No," attach a list. (see instructions)
H(c) Group exemption number ▶

I Tax-exempt status: ☑ 501(c)(3) ☐ 501(c) ( ) ◀ (insert no.) ☐ 4947(a)(1) or ☐ 527

J Website: ▶ COMPASSHEALTHNETWORK.ORG

K Form of organization: ☑ Corporation ☐ Trust ☐ Association ☐ Other ▶

L Year of formation: 1974

M State of legal domicile: MO

Part I **Summary**

Activities & Governance

1 Briefly describe the organization's mission or most significant activities:
TO PROVIDE COMPREHENSIVE, HIGH QUALITY, ACCESSIBLE AND AFFORDABLE HEALTHCARE TO RESIDENTS OF MO & LA WHILE ENCOURAGING THE PERSONAL AND PROFESSIONAL GROWTH OF OUR STAFF.

2 Check this box ▶ ☐

3	Number of voting members of the governing body (Part VI, line 1a)	3	12
4	Number of independent voting members of the governing body (Part VI, line 1b)	4	12
5	Total number of individuals employed in calendar year 2016 (Part V, line 2a)	5	2,597
6	Total number of volunteers (estimate if necessary)	6	12
7a	Total unrelated business revenue from Part VIII, column (C), line 12	7a	0
b	Net unrelated business taxable income from Form 990-T, line 34	7b	0

			Prior Year	Current Year
Revenue	8	Contributions and grants (Part VIII, line 1h)	2,075,285	1,803,876
	9	Program service revenue (Part VIII, line 2g)	132,095,393	143,213,531
	10	Investment income (Part VIII, column (A), lines 3, 4, and 7d)	152,145	414,980
	11	Other revenue (Part VIII, column (A), lines 5, 6d, 8c, 9c, 10c, and 11e)	1,500,712	824,341
	12	Total revenue—add lines 8 through 11 (must equal Part VIII, column (A), line 12)	135,823,535	146,256,728
Expenses	13	Grants and similar amounts paid (Part IX, column (A), lines 1–3)	0	0
	14	Benefits paid to or for members (Part IX, column (A), line 4)	0	0
	15	Salaries, other compensation, employee benefits (Part IX, column (A), lines 5–10)	92,933,679	101,537,973
	16a	Professional fundraising fees (Part IX, column (A), line 11e)	0	0
	b	Total fundraising expenses (Part IX, column (D), line 25) ▶430,365		

Reference: https://www.sec.gov/reportspubs/investor-publications/investorpubsbankrupt

Data Exploration

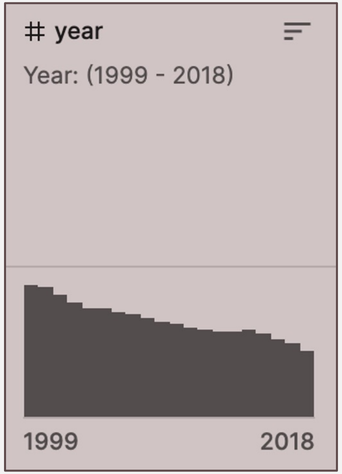
1 Imbalance dataset

Bankrupt (1) 7% 93% Alive (0)

"This dataset contain only 7% unique firms that classified as bankruptcy."

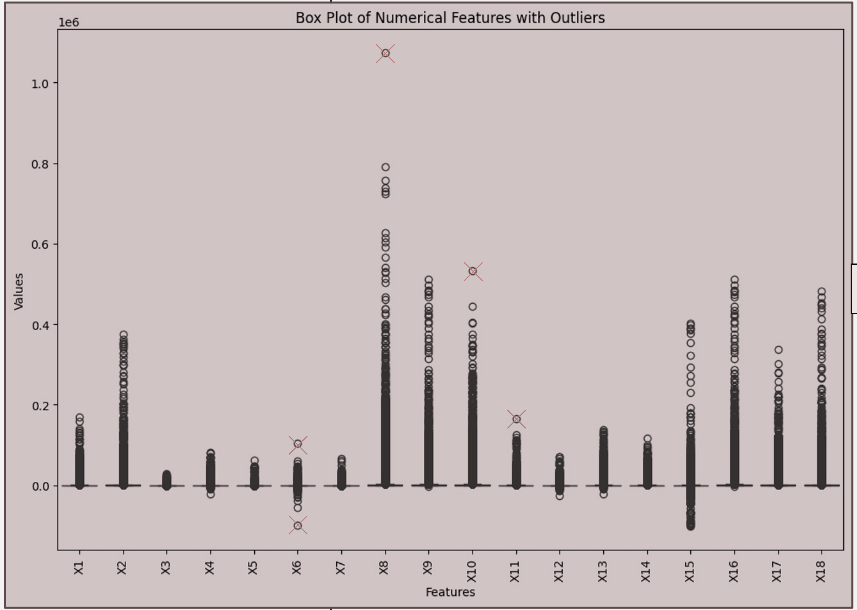
Solving by: Class weight/Resampling

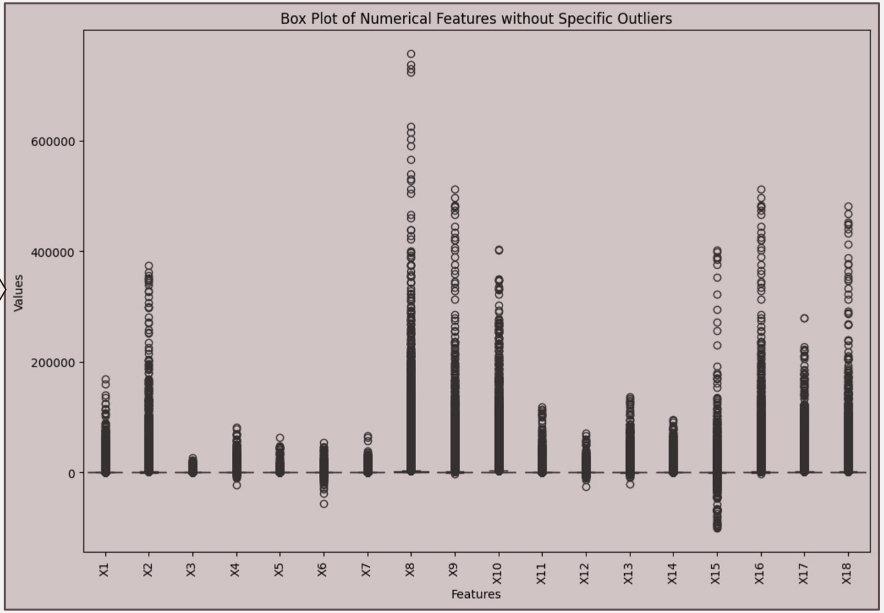
2 Imbalance number of dataset in each period



Data Exploration

Outlier

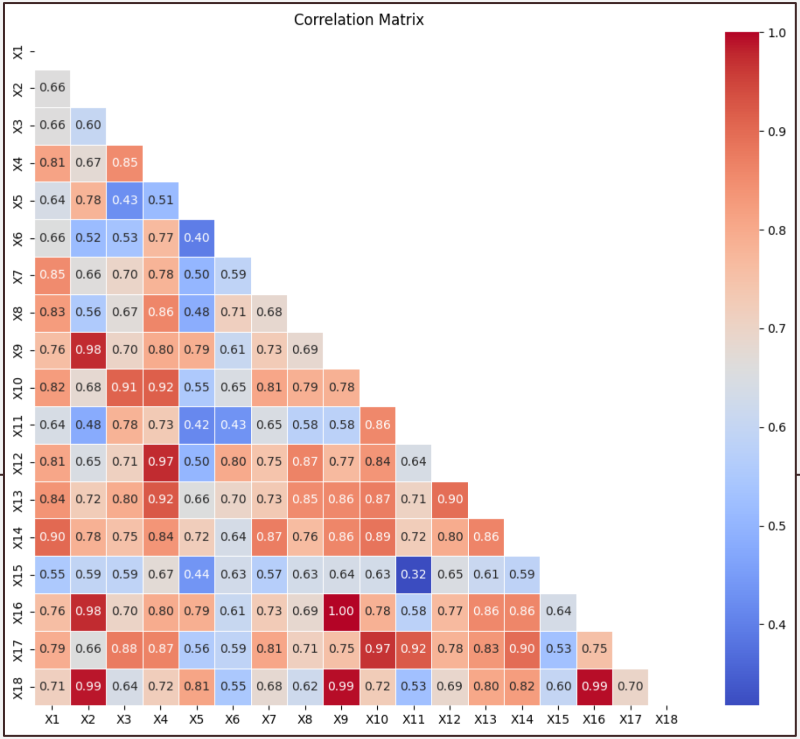




Solving by: Manual Deletion

Data Exploration

Feature Correlation



04

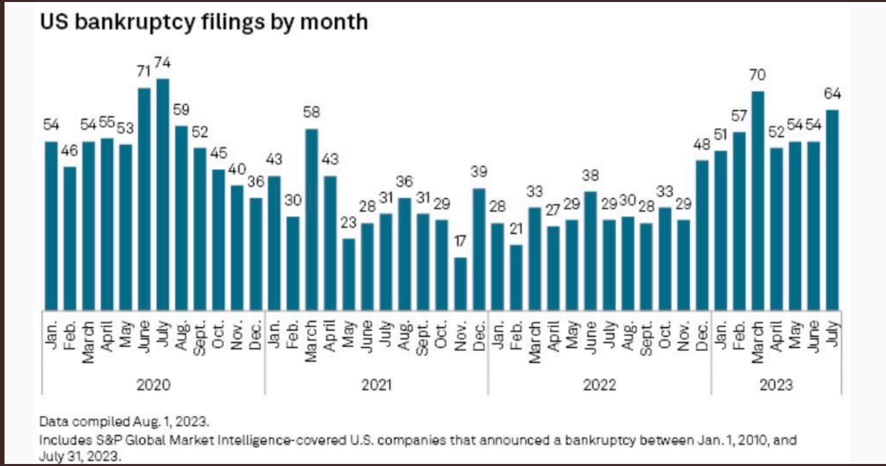
Methodologies

Pipeline of the Project

1. Collect Data
2. Pre-process Data
3. Train Model
4. Evaluate
5. Validate / Showcase

Data Separation

The data have its own story each year, We cannot separate data randomly to train/val/test



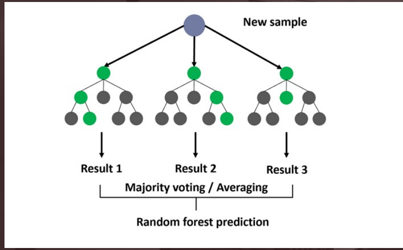

```python
# Training set (-2011) 75%
X_rf_train = df.loc[df["year"] <= 2011, df_rf_selected.columns.drop(["status_label"])]
y_rf_train = df.loc[df["year"] <= 2011, "status_label"]

# Validation set I(2012-2014) 13%
X_rf_test = df.loc[df["year"].between(2012, 2014), df_rf_selected.columns.drop(["status_label"])]
y_rf_test = df.loc[df["year"].between(2012, 2014), "status_label"]

# Test set (2015+) 12%
X_val = df.loc[df["year"] >= 2015, df_rf_selected.columns.drop(["status_label"])]
y_val = df.loc[df["year"] >= 2015, "status_label"]
```

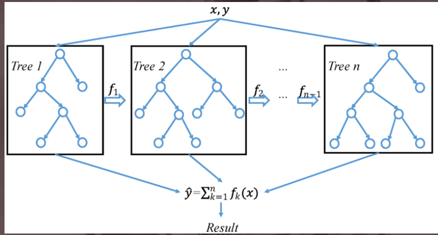
Baseline Models

Random Forest



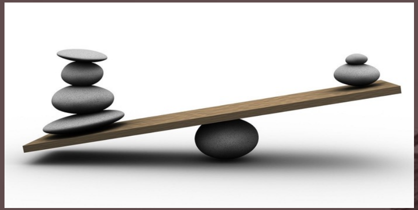
- Ensemble method that combines multiple decision trees to improve accuracy
- Popularity, SOTA model in ML
- Fast to train

XGBoost



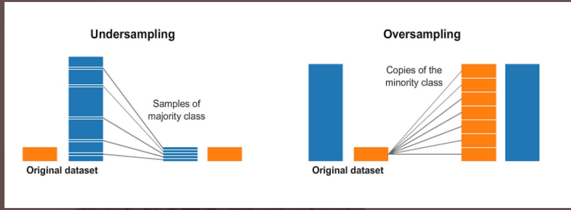
- Ensemble method that combines multiple weaker models to create a stronger one
- gradient boosting can handle nonlinear relationships between features

Implementation of the models: responding to the biased dataset (93% vs 7%)



Class Weighted

Assign higher weights to the minority class, allowing the model to give more importance to its samples during training and reduce bias towards the majority class.



Resampling

Adjust the balance of class distribution in the dataset by oversampling the minority class and undersampling the majority class at the same time. This helps mitigate the impact of imbalanced data, promoting fair representation of both classes during model training.

Training Pipeline

Select all 18 features for
train the models

Random Forest

Set everything default, with a
callback of early stopping = 12

XGBoost

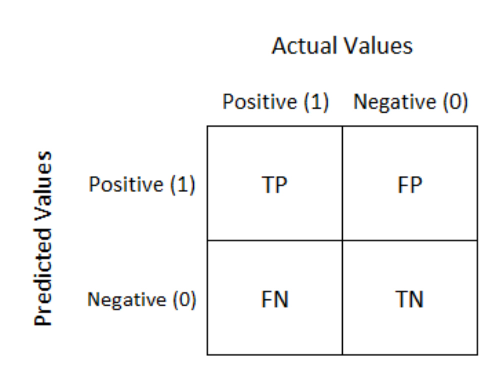
n-estimator = 1,000; with the same callback

With Weighted Method

Increase the positive weight
around 13 times (93/7)

With Resampling Method

use RandomOverSampler
with fixed seed number



Confusion Matrix

Evaluation

Sensitivity: Ability to detect bankruptcy firms

Specificity: Ability to detect healthy firms

Accuracy: Ability to detect firms responding to its classes

F1-Score: Performance of the model

05 Results

Benchmark Results

Detect bankruptcy firms

Detect healthy firms

Model	True Positive	True Negative	False Positive	False Negative	Sensitivity	Specificity	Accuracy	F1-Score
Random Forrest	15	11978	17	272	0.05	1.00	0.98	0.09
XG Boost	35	11975	20	252	0.12	1.00	0.98	0.20
Weighted XG Boost	201	11099	1896	86	0.70	0.85	0.85	0.17
Resampling XG Boost	210	9565	2430	77	0.73	0.80	0.80	0.14

= Highest Value

Our Chosen Model

Model	True Positive	True Negative	False Positive	False Negative	Sensitivity (Detect bankruptcy firms)	Specificity (Detect healthy firms)	Accuracy	F1-Score
Random Forrest	15	11978	17	272	0.05	1.00	0.98	0.09
XG Boost	35	11975	20	252	0.12	1.00	0.98	0.20
Weighted XG Boost	201	11099	1896	86	0.70	0.85	0.85	0.17
Resampling XG Boost	210	9565	2430	77	0.73	0.80	0.80	0.14

06

Conclusions

Showcase#1

HyreCar

BANKRUPTCY

Car-Sharing Platform HyreCar Files for Bankruptcy

Company cites mounting costs and liabilities from lawsuits and investigations, including insider-trading probes by the Justice Department and SEC

By *Becky Yerak*

Feb. 27, 2023 at 2:48 pm ET | WSJ PRO

Share Resize Listen (2 min)

Drivers for ride-hailing and delivery companies can register, apply and extend a vehicle rental through HyreCar's web platform or mobile application. PHOTO: BRENDAN MCDERMID/REUTERS

HyreCar Inc., a publicly traded car-sharing platform, has filed for bankruptcy, partly blaming mounting legal fees from numerous lawsuits and investigations, including Justice Department and Securities and Exchange Commission probes into insider stock trading.

About Company

HyreCar Inc. operates a car-sharing marketplace in the United States. Its marketplace allows car owners to rent their idle cars to ride-sharing service drivers. The company sourcing vehicles from individual owners, as well as commercial owners of vehicles including car dealerships and fleet owners. HyreCar Inc. was incorporated in 2014 and is headquartered in Los Angeles, California.

HyreCar filed a voluntary Chapter 11 petition with the bankruptcy court on February 23.

Our Result

The model can not predict a chance to get bankrupt

company_name	year	True_Label	Predicted_Label
HYREQ	2020	0	0
HYREQ	2021	0	0

Showcase#2

WeWork Files for Bankruptcy

The struggling coworking company filed for Chapter 11 late Monday and plans to exit more leases

BY NICHOLAS RIZZI NOVEMBER 6, 2023 10:07 PM REPRINTS

THE WEWORK LOGO DISPLAYED OUTSIDE AN OFFICE BUILDING IN LOS ANGELES.

PHOTO: PATRICK T. FALLON/AFP VIA GETTY IMAGES

About Company

WeWork Inc. was founded in 2010 and is headquartered in New York. WeWork provides flexible workspace solutions to individuals and organizations worldwide.

On November 6, 2023, WeWork Inc., along with its affiliates, filed a voluntary petition for reorganization under Chapter 11 in the U.S. Bankruptcy Court for the District of New Jersey.

Our Result

The model can predict a chance to get bankrupt so that mean the prediction can warn for investor this company will be bankrupt.

company_name	year	True_Label	Predicted_Label
WeWork	2020	0	1
WeWork	2021	0	1
WeWork	2022	0	1

Showcase#3

About Company

SmileDirectClub was a teledentistry company. The company was co-founded in 2014 by Jordan Katzman and Alex Fenkell. It was based in Nashville, Tennessee, United States.

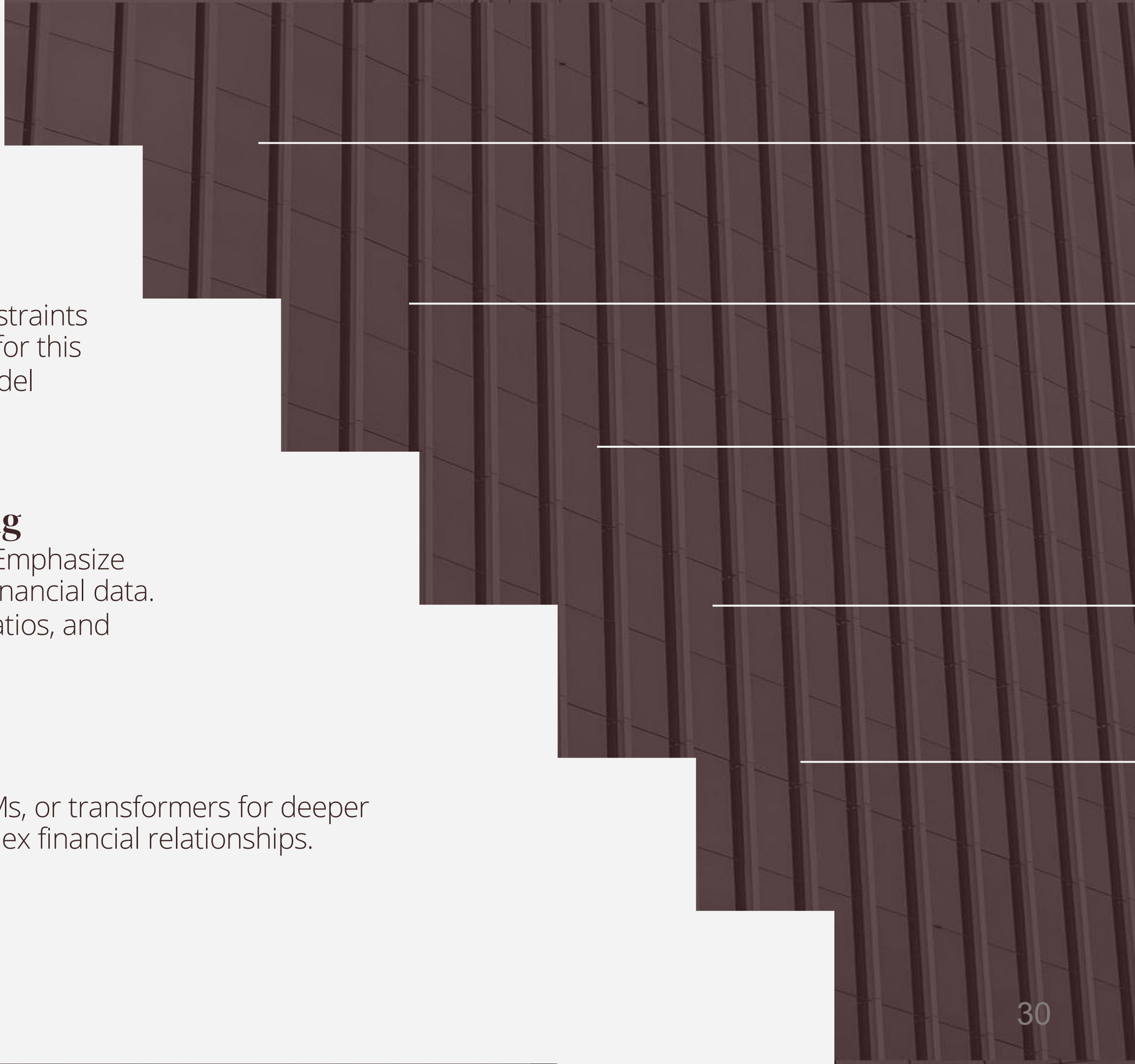
SmileDirectClub shut down in December 2023, less than three months after filing for Chapter 11 bankruptcy.

Our Result

The model can not predict a chance to get bankrupt.

company_name	year	True_Label	Predicted_Label
SmileDirectClub	2020	0	0
SmileDirectClub	2021	0	0
SmileDirectClub	2022	0	0

Lessons Learned



Feature Selection

Feature correlation was insightful, but time constraints hindered feature selection. Allocate more time for this crucial step in future projects for enhanced model efficiency.

Data Quality and Feature Engineering

Optimize results with quality, diverse datasets. Emphasize feature engineering for valuable insights from financial data. Enhance quality by adding financial numbers, ratios, and sector-specific models.

Advanced Architectures

Explore advanced architectures like RNNs, LSTMs, or transformers for deeper insights into temporal dependencies and complex financial relationships.

Recommendations

Hyperparameter Tuning

Optimize model performance through thorough hyperparameter tuning. Experiment with learning rates, batch sizes, and regularization techniques for the best configuration tailored to specific problem.

Ensemble Learning

Boost predictive accuracy and robustness by implementing ensemble learning techniques, like bagging or boosting, which combine predictions from multiple models.

Regular Updates

Ensure model relevance by scheduling regular updates. Retrain periodically with the latest financial data to maintain accuracy as the landscape evolves.

Collaboration with Domain Experts

Collaborate with analysts and domain experts for valuable insights on feature selection and model evaluation. Leverage domain expertise to refine the model, ensuring alignment with real-world financial scenarios.

Q & A

Thank you