

Predicción de Bancarrota en Azure

Cloud Computing (MA3001B)

Desarrollo de proyectos de ingeniería matemática

Gpo 502

Profesor: Jesús Jorge Armenta Segura

Integrantes del equipo

Nombre	Matrícula	Departamento
Kevin Antonio González Díaz	A01338316	Cómputo en la nube
Luis Maximiliano López Ramírez	A00833321	Cómputo en la nube
Gadiel Gerardo Wisar Sotelo	A00834274	Modelos
Julio Eugenio Guevara Galván	A01704733	Modelos
Rodolfo Jesús Cruz Rebollar	A01368326	Datos

[Repositorio Github](#)

4 de abril de 2024

Índice

1. [Objetivo](#)
2. [Requisitos técnicos](#)
3. [Instrucciones de uso](#)
 - [Paso 1: Fork it](#)
 - [Paso 2: Entrenamiento del modelo](#)
 - [Paso 3: Despliega el modelo en Azure](#)
 - [Paso 4: Prueba la API](#)
4. [Archivos del proyecto](#)
5. [URI del servicio](#)
6. [Validación final](#)
7. [Dataset utilizado](#)
8. [Repositorio de Github](#)

Objetivo

Construir un servicio web basado en modelos de machine learning que, al recibir datos financieros de una empresa, pueda predecir si está en riesgo de bancarrota.

Requisitos técnicos

- Python 3.10+
- Azure ML SDK (azureml-core)
- XGBoost
- scikit-learn
- pandas
- requests

Especificaciones Equipo Utilizado

MacBook Air	
M1, 2020	
Chip	Apple M1
Memoria	8 GB
Disco de arranque	Macintosh HD
Número de serie	FVFFMB72Q6L7
macOS	Sequoia 15.3.1

Instrucciones de uso

Paso 1: Fork it

Primero hicimos fork de este repositorio del profe en nuestra máquina local para ejecutarlo y ver el funcionamiento:

```
Unset  
git clone https://github.com/JesusASmx/Cloud_Computing_Class2
```

Paso 2: Entrenamiento del modelo

```
Python  
Model.ipynb
```

Desarrollamos este modelo híbrido XGBoost + Regresión Logística, ajustado con umbral dinámico para balancear recall y precisión de la clase minoritaria.

Archivos generados:

- `model_bankruptcy.pkl`: Es el archivo que contiene el modelo de machine learning ya entrenado
 - `umbral_bankruptcy.json`: Es un archivo que guarda el umbral dinámico (threshold) que se usa para decidir si se entra en clase 0 o 1.
-

Paso 3: Despliegue del modelo en Azure

```
Python
Deployer.ipynb
```

Usando nuestro id de suscripción de Azure registramos y desplegamos el modelo en Azure y generamos el archivo:

- `uri.json`: Es un archivo que contiene la URL del endpoint web donde se desplegó el modelo en Azure.
-

Paso 4: Prueba la API

En `API.ipynb` lo configuramos para enviar datos al servicio web primero con 20 datos aleatorios del dataset utilizado para entrenamiento 10 clasificados como 1 y 10 como 0 y tuvo buen rendimiento. Después lo probamos en `API2.ipynb` con el de prueba que nos compartió, en ese acertó 8 de 10. Ambos archivos están en el repositorio de github.

URI del servicio

```
Python
{"URI":
["http://0ad22dee-50c4-4abd-afde-2831bd2566a4.centralindia.azurecontainer.io
/score"]}
```

Validación final

Probamos el modelo enviando 20 registros aleatorio 10 con Bankrupt? = 1 y 10 con Bankrupt? = 0. El modelo logró distinguir correctamente la mayoría de los casos.

-industry come and s/revenue	Continuous interest rate (after tax)	...	No- credit Interval	Gross Profit to Sales	Net Income to Stockholder's Equity	Liability to Equity	Degree of Financial Leverage (DFL)	Interest Coverage Ratio (Interest expense to EBIT)	Net Income Flag	Equity to Liability	Real	Predicción
0.303810	0.781701	...	0.623269	0.604139	0.841614	0.281304	0.026948	0.565811	1	0.024848	1	1
0.303582	0.781676	...	0.624074	0.605255	0.843489	0.278285	0.026791	0.565158	1	0.037092	0	0
0.303496	0.781648	...	0.623876	0.615920	0.842222	0.285282	0.026870	0.565507	1	0.019363	0	0
0.302249	0.781061	...	0.623173	0.586607	0.916329	0.218785	0.026745	0.564930	1	0.008500	1	1
0.303344	0.781479	...	0.624928	0.599503	0.836943	0.292375	0.026118	0.556329	1	0.015703	1	1
0.302721	0.781055	...	0.623689	0.598911	0.824230	0.292176	0.026697	0.564666	1	0.015765	1	1
0.303451	0.781743	...	0.625407	0.611967	0.840884	0.277676	0.026885	0.565571	1	0.042552	0	0
0.303413	0.781230	...	0.623797	0.609344	0.831255	0.286656	0.026645	0.564353	1	0.018316	1	1
0.303563	0.781636	...	0.624636	0.600301	0.842056	0.280399	0.026852	0.565433	1	0.027162	0	0
0.303400	0.781515	...	0.623918	0.602862	0.838623	0.280054	0.026718	0.564781	1	0.028248	0	0
0.303326	0.781515	...	0.623244	0.598436	0.837267	0.284545	0.026559	0.563737	1	0.020046	0	0
0.302508	0.780830	...	0.623288	0.599360	0.344652	0.643692	0.026781	0.565108	1	0.010494	1	1
0.303521	0.780226	...	0.623419	0.536813	0.787293	0.307630	0.026765	0.565032	1	0.013170	1	1
0.303519	0.781692	...	0.624461	0.605354	0.842509	0.277258	0.026817	0.565280	1	0.047805	0	0
0.303523	0.781488	...	0.622586	0.597308	0.838887	0.282784	0.026125	0.556567	1	0.022179	1	1
0.303481	0.781632	...	0.624203	0.605785	0.840704	0.276773	0.026811	0.565249	1	0.056514	0	0
0.303421	0.781445	...	0.623704	0.611823	0.838304	0.285822	0.026318	0.561074	1	0.018921	1	0
0.303531	0.781645	...	0.623034	0.607520	0.840529	0.281733	0.026791	0.565159	1	0.023960	0	0
0.303401	0.781720	...	0.623888	0.613937	0.843529	0.280201	0.026804	0.565218	1	0.027769	0	0
0.303082	0.781085	...	0.623637	0.585875	0.798728	0.287020	0.026771	0.565059	1	0.018077	1	1

En la parte 2 usando el archivo prueba.csv, el modelo acertó 8 de 10 casos:

No-credit Interval	Gross Profit to Sales	Net Income to Stockholder's Equity	Liability to Equity	Degree of Financial Leverage (DFL)	Interest Coverage Ratio (Interest expense to EBIT)	Net Income Flag	Equity to Liability	Real	Predicción
483.525085	91398.771828	185869.684595	56136.475541	6439.213877	194339.999725	218748	-21751.619260	0	0
459.324081	89736.384811	186528.801063	63877.101621	6446.586494	194383.053127	218748	-30299.884588	0	0
1464.519612	91705.142213	186246.487611	55978.622787	6426.802673	194262.498897	218748	-18209.737872	0	0
1224.732013	88276.137513	185237.572268	58382.352910	6204.241904	187676.849320	218748	-29077.210429	1	0
142.316882	90901.039680	186230.436500	56715.036089	6434.530874	194311.528008	218748	-26386.100504	0	0
408.065819	89383.825463	185875.550424	57709.995625	6441.210105	194351.866011	218748	-28496.495732	0	0
398.005486	90352.769763	185871.470625	56714.641320	6442.092860	194357.063173	218748	-26384.535195	0	0
1241.135760	89792.593103	185154.893963	57580.732969	6396.280314	194039.800334	218748	-28340.347833	1	0
1024.075425	93102.294268	186386.240393	56728.261505	6427.416966	194266.493014	218748	-26437.924626	0	0
1018.642896	91698.156767	185792.157626	56500.961406	6460.832308	194460.613492	218748	-25342.485140	0	0

Dataset utilizado

<https://www.kaggle.com/datasets/fedesoriano/company-bankruptcy-prediction>

Repositorio Github

<https://github.com/JulioGuevaraA01704733/Equipo2ModNube/tree/main>