**PREDICCION DE QUIEBRA EN UNA EMPRESA**

Estudiante: Mamani Quiroga Marcelino

Carrera: Informática

Universidad Mayor de San Andrés

Hoy en dia la competitividad entre empresas de algún rubro es enorme

tanto asi que las herramientas informáticas para acelerar y facilitar los procedimientos de la empresa son importantes para controlar que los objetivos de la empresa se están cumpliendo, es así que la IA también ha tocado este ámbito.

1. **INTRODUCCION**

Las herramientas informáticas basadas en Machine Learning e Inteligencia artificial facilitan en gran manera a la toma de decisiones en una organización cualquiera. A diario las empresas generan grandes cantidades de informacion sobre sus operaciones, clientes, estadísticas y demás. Es justamente esta informacion la que alimenta a una herramienta de IA, desde la cual intenta predecir futuros sucesos. De ese modo presentamos este articulo con un ejemplo de implementación básica de un modelo de aprendizaje alimentado de datos reales obtenidos de una DATASET obtenido de KAGGLE.

1. **RECOPILACION DE DATOS**

Si visitamos la pagina <https://www.kaggle.com/> podemos encontrarnos con miles de datos compartidos por usuarios de todo tipo.

Nos interesa trabajar con datos específicos que sean de nuestro interés y el siguiente fue el elegido:

company-bankruptcy-prediction

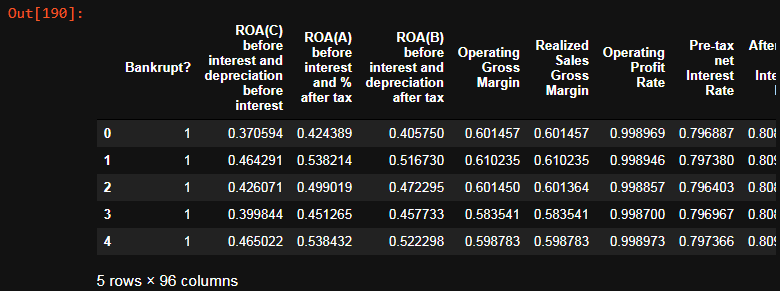
Los datos se obtuvieron del Taiwan Economic Journal para los años 1999 a 2009. La quiebra de la empresa se definió con base en las regulaciones comerciales de la Bolsa de Valores de Taiwán.

Si echamos un vistazo a los datos en la misma pagina veremos las características de cada una:

Y - ¿En quiebra ?: Etiqueta de clase  
X1 - ROA (C) antes de intereses y depreciación antes de intereses: Rentabilidad sobre activos totales (C)  
X2 - ROA (A) antes de intereses y% después de impuestos: Rentabilidad sobre activos totales (A)  
X3 - ROA (B) antes de intereses y depreciación después de impuestos: Retorno sobre activos totales (B)  
X4 - Margen bruto operativo: Ganancia bruta / Ventas netas  
X5 - Ventas realizadas Margen bruto: Ganancia bruta realizada / Ventas netas  
X6 - Tasa de ganancia operativa: Ingresos operativos / Ventas netas  
X7 - Tasa de interés neta antes de impuestos: Ingresos antes de impuestos / Ventas netas  
X8 - Tasa de interés neta después de impuestos: Ingresos netos / Ventas netas  
X9 - Ingresos y gastos / ingresos no industriales: Relación de ingresos netos no operativos  
X10 - Tasa de interés continua (después de impuestos): Ingresos netos - Excluir ganancias o pérdidas por enajenación / Ventas netas  
X11 - Tasa de gastos operativos: Gastos operativos / Ventas netas  
X12 - Tasa de gastos de investigación y desarrollo: (Gastos de investigación y desarrollo) / Ventas netas  
X13 - Tasa de flujo de efectivo: Flujo de efectivo de los pasivos operativos / circulantes  
X14 - Tasa de interés de la deuda que devenga intereses: Deuda que devenga intereses / Patrimonio  
X15 - Tasa de impuestos (A): Tasa de impuestos efectiva  
X16 - Valor neto por acción (B): Valor en libros Por acción (B)  
X17 - Valor neto por acción (A): Valor en libros por acción (A)  
X18 - Valor neto por acción (C): Valor en libros por acción (C)  
X19 - BPA persistente en las últimas cuatro estaciones: BPA -Renta neta  
X20 - Flujo de caja por acción  
X21 - Ingresos por acción (Yuan ¥): Ventas por acción  
X22 - Beneficio operativo por acción (Yuan ¥): Ingresos operativos por acción  
X23 - Beneficio neto por acción antes de impuestos (Yuan ¥): Ingresos antes de impuestos por acción  
X24 - Ventas brutas realizadas Tasa de  
crecimiento de la  
utilidad X25 - Tasa de crecimiento de la utilidad operativa: Crecimiento de la utilidad operativa X26 - Tasa de crecimiento de la utilidad neta después de impuestos: Crecimiento de la  
utilidad neta X27 - Tasa de crecimiento de la utilidad neta regular: Ingresos operativos continuos después del crecimiento de impuestos  
X28 - Tasa de crecimiento de la utilidad neta continua: Ingresos netos -Excluyendo el crecimiento de pérdidas o ganancias por enajenación  
X29 - Tasa de crecimiento de activos totales: Crecimiento de activos totales  
X30 - Tasa de crecimiento de activos netos: Crecimiento de capital total  
X31 - Tasa de crecimiento de retorno de activos totales: Retorno sobre el crecimiento de activos totales  
X32 -% de reinversión en efectivo: índice de reinversión en efectivo  
X33 - índice actual  
X34 - índice rápido: prueba ácida  
X35 - índice de gastos por intereses: gastos por intereses / ingresos totales  
X36 - deuda total / patrimonio neto total: pasivo total / índice de capital  
X37 - índice de deuda% : Pasivo / Activo total  
X38 - Valor neto / Activo: Patrimonio / Activo total  
X39 - Índice de idoneidad del fondo a largo plazo (A): (Pasivo a largo plazo + Patrimonio) / Activo fijo  
X40 - Dependencia del endeudamiento: Costo de la deuda que devenga intereses  
X41 - Pasivos contingentes / Patrimonio neto: Pasivo contingente / Patrimonio  
X42 - Beneficio operativo / Capital pagado: Ingresos operativos / Capital  
X43 - Beneficio neto antes de impuestos / Capital pagado: Ingresos antes de impuestos / Capital  
X44 - Inventario y cuentas por cobrar / Valor neto: (Inventario + Cuentas por cobrar) / Patrimonio  
X45 - Rotación de activos totales  
X46 - Rotación de cuentas por cobrar  
X47 - Días de cobro promedio: Días por cobrar pendientes  
X48 - Tasa de rotación de inventarios (veces)  
X49 - Rotación de activos fijos Frecuencia  
X50 - Tasa de rotación del patrimonio neto (veces):  
Rentabilidad del capital X51 - Ingresos por persona: Ventas por empleado  
X52 - Beneficio operativo por persona: Ingresos operativos por empleado  
X53 - Tasa de asignación por persona: Activos fijos por empleado  
X54 - Capital de trabajo al total Activos  
X55 - Activos rápidos / Activos totales  
X56 - Activos corrientes / Activos totales  
X57 - Efectivo / Activos totales  
X58 - Activos rápidos / Pasivo corriente  
X59 - Efectivo / Pasivo corriente  
X60 - Pasivo corriente a activos  
X61 - Fondos operativos a pasivo  
X62 - Inventario / Capital de trabajo  
X63 - Inventario / Pasivo corriente  
X64 - Pasivo corriente / Pasivo  
X65 - Capital de trabajo / Patrimonio  
X66 - Pasivo corriente / patrimonio  
X67 - Pasivo a largo plazo con activos corrientes  
X68 - Utilidades retenidas con el activo total  
X69 - Ingreso total / gasto  
total  
X70 - Gasto total / Activo X71 - Tasa de rotación del activo corriente: Activo corriente a ventas  
X72 - Rotación rápida de activo Tasa: Activos rápidos a ventas  
X73 - Tasa de rotación del capital de trabajo: Capital de trabajo a ventas  
X74 - Tasa de rotación de efectivo: Efectivo a ventas  
X75 - Flujo de efectivo a ventas  
X76 - Activos fijos a activos  
X77 - Pasivo corriente a pasivo  
X78 - Pasivo corriente a patrimonio  
X79 - Patrimonio a pasivo a largo plazo  
X80 - Flujo de efectivo a activos totales  
X81 - Flujo de efectivo a pasivo  
X82 - CFO a activos  
X83 - Flujo de efectivo a patrimonio  
X84 - Pasivo corriente a activos corrientes  
X85 - Pasivo-activos Marca: 1 si el pasivo total excede los activos totales, 0 en caso contrario  
X86 - Ingresos netos a activos totales  
X87 - Activos totales a precio PNB  
X88 - Intervalo sin crédito  
X89 - Utilidad bruta a ventas  
X90 - Utilidad neta a capital  
contable X91 - Pasivo a capital  
X92 - Grado de apalancamiento financiero (DFL)  
X93 - Coeficiente de cobertura de intereses (Gastos por intereses a EBIT)  
X94 - Marcador de ingresos netos: 1 si los ingresos netos son negativos durante los últimos dos años, 0 en caso contrario  
X95 - Patrimonio a pasivo

Luego de tener claro lo que significa cada uno de los campos, podemos comenzar a indagar de qué forma podemos comenzar a trabajar los datos, asegurarse que no existen datos inválidos, datos faltantes, etc.

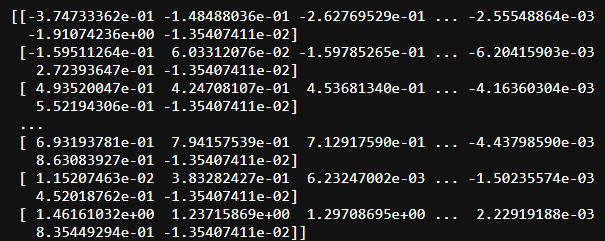
1. **IMPLEMENTACION**



Cada una de las columnas del DATASET son números enteros y reales lo cual facilita el trabajo, sin embargo para construir un modelo de predicción que brinde buenos resultados será necesario un poco más de análisis.

Notemos que la diferencia de valores entre los datos de cada columna varía demasiado, esto puede afectar de gran manera el modelado, es así que deberíamos optar por una forma de escalar estos datos para trabajarlos de mejor manera.

Es importante aclarar que para obtener un modelo que brinde una buena predicción, necesitamos dividir estos datos para el “Entrenamiento” y para “Pruebas”. Esto con el fin de verificar la fiabilidad del modelo desarrollado.



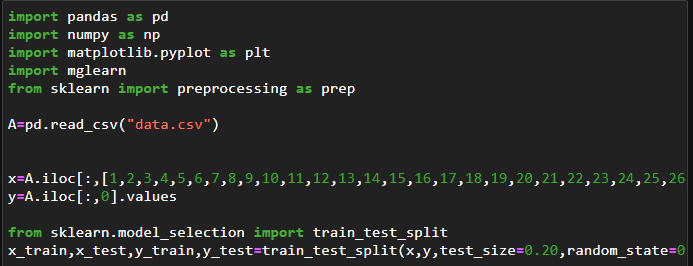
Los datos ya fueron escalados, puesto que son 30 columnas las utilizadas no pueden verse todas, sin embargo con este procedimiento nos aseguramos de que ningún dato no haya sido escalado.

Es momento de elegir un clasificador. La selección de un clasificador depende bastante del tipo de datos que tenemos y de los objetivos de la investigación. El objetivo principal es determinar con una variable binaria si una empresa caerá en bancarrota o no.

**Regresión logística**

La regresión logística es un método estadístico que trata de modelar la probabilidad de una variable cualitativa binaria (dos posibles valores) en función de una o más variables independientes. La principal aplicación de la regresión logística es la creación de modelos de clasificación binaria.

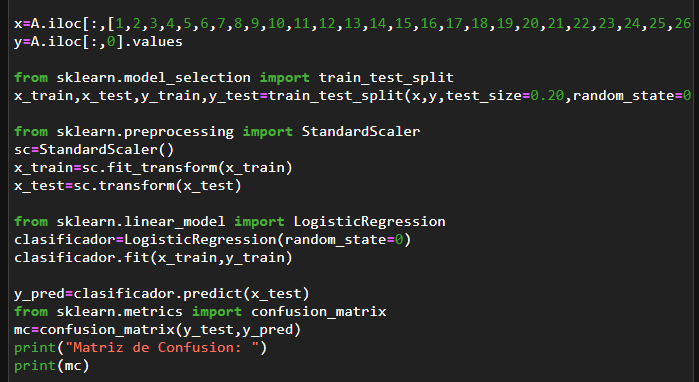
Para el objetivo de esta investigación, el modelo citado se adecua muy bien, sin embargo es posible aplicar otro tipo de clasificadores que tambien podrían brindar buenos resultados.



Hemos importado las librerias correspondientes para preprocesamiento y otras para funciones matemáticas.

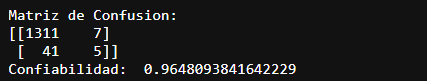
También importamos librerias de SKLEARN para hacer selección de un grupo de datos de entrenamiento y otros de prueba con el fin de realizar evaluaciones al modelo una vez que se realizó el aprendizaje.

Luego de realizar las importaciones de librerias correspondientes a modelos de regresión logistica para Python, procedemos a entrenar el clasificador creado: de esa forma todos los ejemplos seleccionados para entrenamiento entraran al clasificador y podremos hacer prueba de su aprendizaje con los datos de prueba.





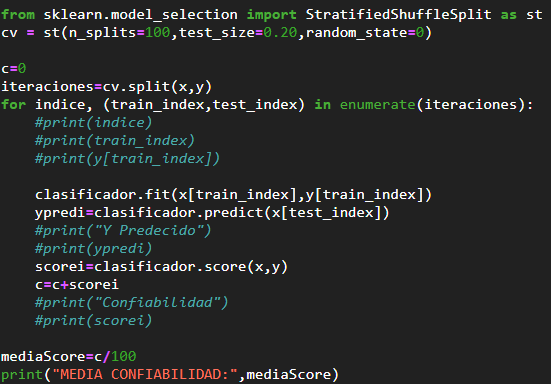
Nuestro clasificador ha sido entrenado, y luego testeado con los datos de prueba, para determinar el nivel de aprendizaje podemos utilizar la llama “Matriz de Confusion” el cual hace un recuento de los aciertos y fallos en la predicción del modelo. Veamos:



Puesto que nuestra variable y solo toma valores de 0 y 1, la matriz de confusión nos indica que para el valor de 0 tuvo 1316 sumando la diagonal principal y unos errores de 48 si sumamos la otra diagonal. Con esto podemos decir que el aprendizaje ha sido bueno. A su vez podemos verificar el nivel de confiabilidad del modelo igual a un 96.4% indicándonos que es es bastante confiable.

Luego de haber realizado una primera prueba con una sola división de los datos, podemos proceder a realizar las divisiones que consideremos necesarias para ver si el aprendizaje puede mejorar.

Lo conseguiremos de la forma siguiente:



Luego de importar las librerias necesarias creamos una variable cv el cual tendrá el numero de entrenamientos o bien, las veces que va a dividir los datos de entrenamiento entre si de forma aleatoria para obtener si es posible, un mejor aprendizaje.

Una vez que las 100 iteraciones se realizaron y nuestro clasificador fue entrenado podemos visualizar el nivel de confiabilidad promedio obtenido de todas las iteraciones.



El nivel de confiabilidad es ligeramente superior al hallado en la primera prueba, esto nos da a entender que entre más entrenamientos tuvo el clasificador, el aprendizaje fue mejorando.

De esta forma podemos comprobar que el clasificador brinda buenos resultados.

1. **CONCLUSIONES**

Es posible realizar un estudio mucho más extenso, utilizando todas las columnas del DATASET y probando diferentes clasificadores para ver el rendimiento y nivel de aprendizaje de cada uno de ellos, sin embargo el modelo utilizado en este artículo nos brinda un punto de arranque sólido para realizar nuevas pruebas puesto que el nivel de confiabilidad tiene un nivel muy aceptable. Aun asi queda un largo camino que recorrer.