

Redes Neuronales aplicados préstamos Peer-to-Peer

Bitzer Arotoma Bacilio
Facultad de Ciencias
Ciencias de la Computación
Lima, Perú
bitzer.arotoma.b@uni.pe

Abstract— El presente documento está relacionado al mercado de préstamos en línea conocido como Peer-to-Peer. Este tipo de préstamo presenta las tasas de interés más bajas a comparación de los bancos al tratarse de una conexión indirecta entre el prestatario con el prestamista y también por poseer un alto riesgo crediticio, por lo que se propone desarrollar un modelo de clasificación para evaluar a los deudores, el modelo de clasificación que se plantea es el de Redes Neuronales.

Keywords—Peer-to-Peer, KNN, factor FICO, AUC.

I. INTRODUCTION (HEADING 1)

El campo de “Análisis de Datos” hoy en día está abarcando varios campos de investigación tanto en campos de la ciencia como en las industrias. Este trabajo tiene mayor interés en el área de la industria bancaria debido a que la data presente que se analizará contiene los datos completos de préstamos para emitidos desde el año 2007 al 2015, el cual es una data muy importante para los bancos ya que se puede evaluar a los prestatarios con un indicador de riesgo para futuros préstamos.

Se plantea desarrollar un modelo que clasifique a los deudores analizando las características principales que tienen estos. La data es brindada por “Lending club” que es un mercado en línea que conecta a los prestatarios y prestamistas, este mercado puede tener las tasas de interés bajas pero el riesgo elevado, por lo que no es muy confiable ni seguro para muchas personas debido a la asimetría que existe al conectar indirectamente al prestatario con el prestamista.

El artículo está organizado de la siguiente forma: en la siguiente sección se mostrará un breve resumen de trabajos relacionados que usan algoritmos de machine learning para crear modelos de clasificación para el análisis de riesgo crediticio, en la sección III se presenta el diseño del experimento donde se explicará las técnicas que se utilizarán, cuales son los objetivos y funciones que se desea obtener para nuestros posteriores resultados.

II. ESTADO DE ARTE

En esta sección se presentan trabajos de investigación ya desarrollados anteriormente que tienen relación con el tema que se desea desarrollar, se explicará esencialmente los algoritmos de machine learning aplicados a la data, estos algoritmos que son de clasificación.

A. Evaluación de riesgos en préstamos sociales mediante random forest

En el presente artículo se desarrolló métodos de machine learning para la data de “Lending Club”, el principal objetivo de este artículo es mostrar que los indicadores financieros tradicionales, como el factor FICO, no son necesariamente buenos indicadores a comparación de la técnica de Random Forest que obtuvo mejores resultados de clasificación.

El trabajo de preprocesamiento y manipulación de datos se realizó con R mientras que para el algoritmo de aprendizaje se usó el software Weka. Se aplicó los siguientes algoritmos K-Nearest neighbors(KNN), Logistic Regression, Support Vector Machine y Random Forest. Para posteriormente llegar a la conclusión que el Random Forest posee la precisión más alta en 78% y el AUC más alto en 0.71 con el RMSE más bajo en 0.42 a comparación de los otros algoritmos[1]. Sin embargo al tener la precisión más alta de clasificar al buen prestatario presenta el problema de clasificar a algunos prestatarios como malos.

TABLE I. COMPARACIÓN DE RENDIMIENTO DE LOS CLASIFICADORES

Rank	Classifier	Accuracy (%)	AUC	RMSE
1	RF	78	0.71	0.42
2	KNN	70.1	0.53	0.55
3	SVM	63.3	0.62	0.68
4	LR	54.5	0.68	0.51

a.

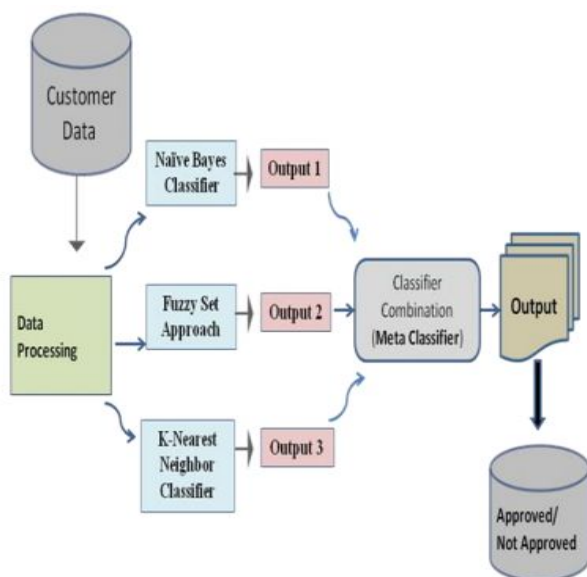
B. Combinación de Random Forest y Redes Neuronales en préstamos Sociales

La finalidad de este artículo es disminuir el riesgo de pérdida al momento de hacer préstamos para lograr esto desarrollaron propusieron un método que consiste en la combinación de Random Forest y Redes Neuronales. En la etapa de preprocesamiento se llegó a la conclusión de evaluar solo 13 de 85 características. La combinación es de evaluar los datos con uno de estos algoritmos para posteriormente entrenar los datos actualizados con uno de estos algoritmos, el que obtuvo mejores resultados fue la combinación en el orden Redes Neuronales para la actualización de los datos y para el entrenamiento usando Random Forest el resultado fue de 73.5% a diferencia de la combinación opuesta que fue de 72%[2], que son dos resultados buenos en comparación de solo aplicar solamente uno de estos algoritmos para el entrenamiento

C. Aprobación de la solicitud de préstamo usando Meta Classifier

El objetivo es diseñar un “Meta Classifier”, el cual consiste en la combinación de Clasificar de Naive Bayes, KNN y Fuzzy Set, al igual que los anteriores artículos mencionados la clasificación consiste en reducir el riesgo crediticio.

El proceso que se usa en este artículo consiste en que el Meta Classifier toma la salida de los clasificadores mencionados como sus entradas para evaluar por mayoría si se aprueba o se rechaza el préstamo, esto consiste básicamente en que si dos clasificadores como mínimo aprobaron el préstamo el Meta Classifier tendrá como salida final aprobado caso contrario rechazará el préstamo. En el siguiente gráfico se muestra todo el flujo del modelo de clasificación[3].



III. DISEÑO DEL EXPERIMENTO

Para desarrollar el modelo de clasificación, se realizará esencialmente en tres etapas, la inicial consistiría en la descripción de la data, así como la evaluación de los valores nulos o datos faltantes, la segunda etapa será el preprocesamiento de la data la cual consiste en la normalización, escalamiento, estandarización y binarización de la data todo esto dependiendo de la característica de la data y el tratado que requiere y por último se aplicará el algoritmo de “redes neuronales” para clasificar la data y obtener nuestro modelo, todo esto se desarrollara con librerías de python como pandas, numpy, scipy, scikit learn, matplotlib, entre otros paquetes.

REFERENCIAS

- [1] Milad Malekipirbazari, Vural Aksakalli, “Risk assessment in social lending via random forests”, Expert Systems with Applications, Volume 42, Issue 10, 2015, Pages 4621-4631.
- [2] Fu, Y. (2017) Combination of Random Forest and Neural Networks in Social Lending. Journal of Financial Risk Management, 6.
- [3] B. Rohit, K. P. Suppreethi, “Loan Application Approval using Meta Classifier”, Dept. of Computer Science & Engineering, JNTUH College of Engineering, Hyderabad.