Nombre: Credit Risk Assessment using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications

Abstract: Una estimación precisa del riesgo y su uso en modelos corporativos o globales de riesgo financiero, podría traducirse en un uso más eficiente de los recursos. Se hace un análisis comparativo de diferentes métodos de satistical y machine learning para clasificación en un conjunto de datos de préstamos hipotecarios con la motivación de comprender sus limitaciones y potencial.

Palabras Clave: Machine Learning, Deep Learning, hipotecas.

Presentación del Problema: Surge el interés de la evaluación del reisgo de crédito a partir de las crisis financieras que tuvieron lugar en las décadas de los 80´s y 90´s. Por lo que una mejor estimación de la precisión del riesgo podría ser traducida en un uso eficiente de los recursos. Entonces la idea crucial de esta investigación radica en encontrar predictores realistas del riesgo individual a través de modelos de riesgo realistas y precisos.

Presentación de la Metodología: La metodología consiste en hacer un análisis comparativo de diferentes métodos de clasificación en un conjunto de datos de préstamos hipotecarios de un gran banco comercial. Lo anterior se realiza mediante un estudio sistemático y una comparación con las técnicas tradicionales de clasificación estadística.

Se utiliza un enfoque de estrategia múltiple donde se aplican varios algoritmos con los mismos datos y se comparan sus resultados para encontrar el mejor modelo. Esto se justifica por el hecho de que es muy difícil seleccionar un modelo óptimo a priori sin conocer la complejidad real de un problema o conjunto de datos en particular. También se introduce una metodología específica para el análisis del modelo basada en el estudio de curvas de error para estimar el ruido / sesgo y la complejidad del modelo y el conjunto de datos.

El proceso o la construcción de un modelo y su aplicación a nuevos ejemplos de datos implican un costo computacional práctico. Esto debe tenerse en cuenta, ya que puede limitar el tipo o los modelos que se pueden utilizar en una situación particular.

Se describen los elementos básicos de la metodología y el análisis de construcción de modelos que se emplean en los cuatro algoritmos considerados en el estudio. Los principales elementos de la metodología de análisis son:

· Exploración de parámetros del modelo básico.

· Análisis de importancia / sensibilidad de las variables.

· Análisis de errores de capacitación, prueba / generalización y evaluación, incluidas las matrices de rendimiento.

· Análisis de curvas de aprendizaje y estimaciones de ruido y parámetros de complejidad.

· Selección de modelo y combinación de resultados.

El CNBV ya había utilizando los datos para un modelo de regresión y, por lo tanto, requería poco procesamiento previo o manipulación antes de la construcción del modelo. Los datos consisten en un único conjunto de datos de 4.000 registros, cada uno de ellos correspondiente a una cuenta de cliente, y contiene un total de 24 atributos.

Los modelos logit y probit son similares, pero utilizan las distribuciones logísticas y normales acumulativas respectivamente. Una diferencia en estas distribuciones es que la distribución logística tiene colas más gruesas y esto a su vez produce pequeñas diferencias en el modelo, sin embargo, no hay bases teóricas para favorecer una técnica sobre la otra.

El modelo CART de Decision-Tree consiste en potentes modelos no paramétricos que producen predicciones precisas y reglas fácilmente interpretables para caracterizarlos. Son buenos representantes de la clase de algoritmos basados en reglas del árbol de decisiones.

Se eligio utilizar la arquitectura de red neuronal feedforward estándar compatible con el conjunto de herramientas de Darwin y se experimento con varios algoritmos de entrenamiento: backpropagation, steepest descent, conjugate gradient, modified Newton, y genetic algorithm Los métodos de segundo orden, como el gradiente conjugado, permiten un entrenamiento mucho más rápido que la retropropagación estándar. También se investigo el efecto de cambiar las funciones de activación para la capa oculta: sigmoide, lineal e hipertangente. El algoritmo genético permite la optimización del peso en la región de la superficie de error, lo que podría ser difícil para los métodos basados en gradientes.

Resultados: Los resultados muestran que los modelos de árbol de decisión CART proporcionan la mejor estimación de incumplimiento con una tasa de error promedio de 8.31% para una muestra de entrenamiento de 2,000 registros. Como resultado del análisis de la curva de error para este modelo, concluimos que si hubiera más datos disponibles, aproximadamente 22,000 registros, se podría lograr una tasa de error potencial de 7.32%. Neural Networks proporcionó el segundo mejor resultado con un error promedio de 11.00%. El algoritmo K-Nearest Neighbour tenía una tasa de error promedio de 14.95%. Estos resultados superaron al algoritmo Probit estándar que alcanzó una tasa de error promedio del 15,13%.

Conclusiones: Si bien el articulo es algo longevo contribuye a la ciencia por que experimenta con varios modelos para un problema que sigue vigente hoy en dia, el problema de predecir si un clinte me pudede pagar el dinero que le preste. La combinación de diferentes estrategias y la aplicación de una metodología sistemática de construcción y selección de modelos ofrecen una perspectiva interesante para comprender las características y la utilidad de diferentes algoritmos o métodos de ajuste de datos

\*\*

Nombre: Deep learning with long short-term memory networks for financial market predictions

Abstract

Palabras Clave

Presentación del Problema: el problema es…

Presentación de la Metodología: La metodología consiste en…

Resultados: El resultado es…

Conclusiones: Contribuye a la ciencia porque…

\*\*

Nombre: Deep Learning for Mortgage Risk

Abstract

Palabras Clave

Presentación del Problema: el problema es…

Presentación de la Metodología: La metodología consiste en…

Resultados: El resultado es…

Conclusiones: Contribuye a la ciencia porque…