

Segmentación y Clasificación de Aprobaciones de Tarjetas



Contexto

Las tarjetas de puntuación de crédito son un método común de control de riesgos en la industria financiera. Utiliza la información personal y los datos presentados por los solicitantes de tarjetas de crédito para predecir la probabilidad de futuros incumplimientos y préstamos de tarjetas de crédito. El banco puede decidir si emite una tarjeta de crédito al solicitante. Los puntajes de crédito pueden cuantificar objetivamente la magnitud del riesgo.

En términos generales, las tarjetas de puntuación de crédito se basan en datos históricos. Una vez que se encuentra con grandes fluctuaciones económicas. Los modelos anteriores pueden perder su poder predictivo original. El modelo logístico es un método común para la calificación crediticia. Porque Logistic es adecuado para tareas de clasificación binaria y puede calcular los coeficientes de cada característica. Para facilitar la comprensión y el funcionamiento, la tarjeta de puntuación multiplicará el coeficiente de regresión logística por un valor determinado (como 100) y lo redondeará.

En la actualidad, con el desarrollo de algoritmos de aprendizaje automático. Se han introducido métodos más predictivos como Boosting, Random Forest y Support Vector Machines en la calificación de tarjetas de crédito. Sin embargo, estos métodos a menudo no tienen buena transparencia. Puede ser difícil proporcionar a los clientes y reguladores una razón para el rechazo o la aceptación.

Tarea

Cree un modelo de aprendizaje automático para predecir si un solicitante es un cliente "bueno" o "malo", a diferencia de otras tareas, no se proporciona la definición de "bueno" o "malo". Debe usar alguna técnica de segmentación para construir su etiqueta (considerar el perfil a la hora de definir si es bueno o malo – sustentarlo, ya

que el modelo de segmentación dará solo clústers). Si es necesario se puede incorporar una etiqueta adicional bajo sustento (bueno, intermedio y malo), esta también debe ser hecho por el modelo de segmentación.

Si se desea, en vez de usar clustering, se pueden usar métodos genéricos, según alguna variable que infiera que es bueno o malo (cuidar que si es variable causal no usarla en el modelo de clasificación).

Tomar cierto grupo de datos para testeo desde el archivo data.csv.