Conclusão

O modelo fez um bom trabalho no geral. Foi preciso na maioria das vezes, acertando em torno de 87,68% dos casos.

Foi particularmente bom em dizer “sim” às pessoas certas, com cerca de 90,31% das respostas “sim” corretas.

Também detectou muitos dos casos reais de “sim”, cerca de 96,12%. No entanto, teve um pouco de dificuldade em dizer “não” corretamente,

com apenas 32,15% de suas respostas “não” acertadas. A pontuação geral, chamada de F1, foi de 93,12%, mostrando que se saiu bem em equilibrar tudo.

É importante lembrar que o conjunto de dados não foi balanceado, pois 87% dos clientes foram classificados como bons pagadores,

enquanto apenas 13% foram considerados mau perfil. Além disso, o conjunto de dados continha apenas pessoas que realmente obtiveram o cartão,

e não aquelas que foram inicialmente negadas por outros motivos, como dívidas não pagas. Isto pode afetar o desempenho do modelo em situações reais.

No futuro, seria uma boa ideia melhorar a capacidade do modelo de dizer “não” com mais precisão e considerar estes desequilíbrios,

ao mesmo tempo que melhora o modelo.

Conclusão

Categoria de clientes que devem ser visados:

Na faixa etária de 30 a 50 anos.

Clientes que trabalham há mais de 10 anos.

Clientes mulheres trabalhadoras.

Clientes com formação académica.

Clientes repetidos.

Conclusões

Nosso modelo foi desenvolvido com base em uma série de suposições. Primeiramente, desenvolvemos nossa própria métrica para chegar à variável alvo.

Uma instituição financeira que deseje implementar nosso modelo para ajudar no processo de aprovação de cartão de crédito

pode desenvolver sua própria métrica para categorizar o que constitui um solicitante "Bom" e "Mau".

A seguir, também formulamos nosso problema como querendo minimizar a taxa de verdadeiros negativos (TNR) e,

ao mesmo tempo, maximizar a taxa de falsos negativos (FNR). Num cenário prático, a formulação do problema

pode ser diferente com base no montante dos lucros obtidos pela instituição financeira. Eles podem querer desconsiderar

o FNR porque estão simplesmente interessados ​​em identificar todos os aplicativos “ruins”. Eles terão, portanto, que ajustar

os hiperparâmetros do nosso modelo.

Conclusão

Neste notebook, realizamos diferentes engenharia de recursos, transformações e métodos de sobreamostragem/redução de amostragem para treinar um modelo para classificar créditos bons/ruins com o conjunto de dados desequilibrado fornecido.

Para a seleção do modo, usamos a pontuação f05 como principal critério para lidar com esses dados desequilibrados.

Dos 4 modelos que tentamos treinar acima, Regressão Logística, KNN e RandomForest produzem resultados semelhantes para validação cruzada. Random Forest tem pontuação f05 mais baixa do que outras. Para potenciais necessidades futuras de inferência, selecionamos a Regressão Logística ajustada como modelo final. Em seguida, executamos a Regressão Logística com o conjunto de teste e ela produz um resultado semelhante à sua pontuação CV f05, que é cerca de 0,8.

A transformação WOE não apresentou muito efeito na Regressão Logística, provavelmente porque realizamos transformação apenas em variáveis ​​categóricas com cardinalidade inferior a 20, deixando de fora aquelas com alta cardinalidade e variável contínua.

Concluindo, acreditamos que nosso projeto atingiu os objetivos definidos em nossa Formulação de Problemas e que serve de base para o desenvolvimento de modelos mais complexos para enfrentar esta questão de Aprovação de Cartão de Crédito.

Conclusão

Randomforest com subamostragem fornece o melhor resultado para encontrar os dados minoritários. Com valor de recall de 0,67 para encontrar os clientes rejeitados é o maior deste projeto. Se o objetivo do banco é encontrar quem rejeitar, este modelo pode ser utilizado. No entanto, se o custo da rejeição incorrecta do cliente for demasiado elevado, outros métodos podem ser utilizados. Por exemplo, o Classificador de Árvore de Decisão sem reamostragem tem taxas de precisão mais altas. Devido à natureza desequilibrada dos dados, permite aceitar mais pedidos de cartão de crédito dos clientes. Com este modelo haverá mais clientes sendo aceitos para cartão de crédito, enquanto com randomforest com subamostragem haverá mais clientes rejeitados corretamente. Na fase de implementação estes resultados devem ser considerados.

Conclusão¶

O modelo foi fantástico na identificação de quais clientes aprovar, tendo precisão de 0,98 e pontuação f1 de 0,99, quase perfeita. No entanto, não foi bom em identificar quais clientes negar, tendo apenas uma precisão de 0,23 e uma pontuação f1 de 0,13. Isso pode ter surgido da falta de clientes negados no conjunto de dados, considerando que havia apenas 300, enquanto havia 14.693 clientes aprovados, tornando o modelo muito melhor na identificação de bons clientes. O modelo em si parece funcionar bem, mas, para ser utilizado de forma eficaz, seria necessário melhorar a capacidade de negar clientes, o que exigiria uma amostra maior de maus clientes com os quais aprender.

Conclusões

Nosso modelo foi desenvolvido com base em uma série de suposições. Primeiramente, desenvolvemos nossa própria métrica para chegar à variável alvo. Uma instituição financeira que deseje implementar nosso modelo para ajudar no processo de aprovação de cartão de crédito pode desenvolver sua própria métrica para categorizar o que constitui um solicitante "Bom" e "Mau".

A seguir, também formulamos nosso problema como querendo minimizar a taxa de verdadeiros negativos (TNR) e, ao mesmo tempo, maximizar a taxa de falsos negativos (FNR). Num cenário prático, a formulação do problema pode ser diferente com base no montante dos lucros obtidos pela instituição financeira. Eles podem querer desconsiderar o FNR porque estão simplesmente interessados ​​em identificar todos os aplicativos “ruins”. Eles terão, portanto, que ajustar os hiperparâmetros do nosso modelo.

Concluindo, acreditamos que nosso projeto atingiu os objetivos definidos em nossa Formulação de Problemas e que serve de base para o desenvolvimento de modelos mais complexos para enfrentar esta questão de Aprovação de Cartão de Crédito.

A pontuação de crédito é uma técnica de gestão de risco amplamente utilizada no setor financeiro. Utiliza informações e dados pessoais fornecidos pelos solicitantes de cartão de crédito para estimar futuras falências e empréstimos de cartão de crédito. O banco tem autoridade para determinar se deve ou não fornecer um cartão de crédito ao requerente. As pontuações de crédito podem estimar o nível de risco objetivamente.

👉 Um modelo de aprendizado de máquina é necessário para prever se um aplicativo é um usuário 'bom' ou 'ruim' neste conjunto de dados. No entanto, não existe uma definição de “bom” ou “mau”. Além disso, o problema de dados desequilibrados é um problema significativo neste projeto.

Os cartões de pontuação de crédito são um método comum de controle de risco no setor financeiro. Ele usa informações e dados pessoais enviados por solicitantes de cartão de crédito para prever a probabilidade de futuras inadimplências e empréstimos de cartão de crédito. O banco pode decidir se deve emitir um cartão de crédito ao requerente. As pontuações de crédito podem quantificar objetivamente a magnitude do risco.

O departamento de cartão de crédito de um banco é um dos principais adotantes da ciência de dados. A aquisição de novos usuários de cartão de crédito é sempre uma prioridade fundamental para o banco. Fornecer cartões de crédito sem a devida diligência ou avaliação da solvabilidade é um risco enorme.

Nas últimas décadas, o departamento de cartão de crédito tem usado uma metodologia de avaliação de crédito baseada em dados chamada Credit Scoring, e o modelo é chamado de scorecard de aplicação. O scorecard de aplicação auxilia no cálculo do nível de risco associado a um solicitante e com base na prioridade estratégica de cada momento, eles decidem o valor de corte da pontuação para aprovação ou rejeição de uma solicitação de cartão de crédito.

Os cartões de pontuação de crédito são um método comum de controle de risco no setor financeiro. Ele usa informações e dados pessoais enviados por solicitantes de cartão de crédito para prever a probabilidade de futuras inadimplências e empréstimos de cartão de crédito. O banco pode decidir se deve emitir um cartão de crédito ao requerente. As pontuações de crédito podem quantificar objetivamente a magnitude do risco.  
   
De modo geral, os cartões de pontuação de crédito são baseados em dados históricos. Depois de encontrar grandes flutuações econômicas. Os modelos anteriores podem perder o seu poder preditivo original. O modelo logístico é um método comum para pontuação de crédito. Porque Logística é adequada para tarefas de classificação binária e pode calcular os coeficientes de cada recurso. Para facilitar a compreensão e a operação, o cartão de pontuação multiplicará o coeficiente de regressão logística por um determinado valor (como 100) e o arredondará.  
   
Atualmente, com o desenvolvimento de algoritmos de aprendizado de máquina. Métodos mais preditivos, como Boosting, Random Forest e Support Vector Machines, foram introduzidos na pontuação de cartão de crédito. No entanto, estes métodos muitas vezes não têm boa transparência. Pode ser difícil fornecer aos clientes e reguladores um motivo para rejeição ou aceitação.

**Tarefa**

Construir um modelo de aprendizado de máquina para prever se um candidato é um cliente 'bom' ou 'ruim', diferente de outras tarefas, a definição de 'bom' ou 'ruim' não é dada. Você deve usar alguma técnica, como [análise vintage,](https://www.kaggle.com/rikdifos/eda-vintage-analysis) para construir seu rótulo. Além disso, o problema de desequilíbrio de dados é um grande problema nesta tarefa.

Cartão de crédito:-¶

Um cartão de crédito é um pedaço retangular fino de plástico ou metal emitido por um banco ou empresa de serviços financeiros que permite aos titulares do cartão emprestar fundos para pagar bens e serviços com comerciantes que aceitam cartões como forma de pagamento. Os cartões de crédito impõem a condição de que os titulares do cartão devolvam o dinheiro emprestado, acrescido de quaisquer juros aplicáveis, bem como quaisquer encargos adicionais acordados, integralmente até a data de cobrança ou ao longo do tempo.

Vantagens do cartão de crédito: -

Ganhe recompensas como dinheiro de volta ou pontos de milhas.

Proteção contra fraudes com cartão de crédito.

Informações de pontuação de crédito gratuitamente.

Sem taxas de transação estrangeira.

Aumento do poder de compra.

Não vinculado a conta corrente ou poupança.

Reter um carro alugado ou um quarto de hotel.

Elegibilidade para cartão de crédito: -

Idade

Renda

Residência

Cidadania

Por que a sua proposta é importante no mundo de hoje? Até que ponto prever um bom cliente vale para um banco?

A) No mundo de hoje, onde as transações online se tornaram a norma, a fraude com cartões de crédito tornou-se uma grande preocupação para bancos e instituições financeiras. É essencial que os bancos prevejam com precisão se um cliente apresenta um risco de crédito bom ou mau para minimizar o risco de perdas financeiras por incumprimento ou fraude. É aqui que entra a previsão de aprovação do cartão de crédito.

Como isso vai impactar o setor bancário?

Ao analisar o histórico de crédito, a renda, a situação profissional e outros dados relevantes de um cliente, os bancos podem prever a probabilidade de um cliente deixar de pagar com cartão de crédito. Isso pode ajudar o banco a tomar decisões informadas sobre aprovar ou não um pedido de cartão de crédito e qual limite de crédito atribuir.

Qual é a lacuna no conhecimento ou como o método proposto pode ser útil?

Prever um bom cliente é importante para um banco por vários motivos.

Ajuda o banco a minimizar o risco de perdas financeiras devido a inadimplência ou fraude.

\*Ajuda o banco a determinar um limite de crédito apropriado para o cliente, o que pode ajudar a evitar que o cliente acumule dívidas que não possa pagar.

Ajuda o banco a identificar potenciais clientes de alto valor que provavelmente gastarão mais em seus cartões de crédito e gerarão mais receitas para o banco.

O departamento de cartão de crédito de um banco é um dos principais adotantes da ciência de dados. Um dos principais focos do banco sempre foi a aquisição de novos clientes de cartão de crédito. Distribuir cartões de crédito sem fazer uma pesquisa adequada ou avaliar a qualidade de crédito dos candidatos é bastante arriscado. O departamento de cartão de crédito usa há muitos anos um sistema baseado em dados para avaliação de crédito chamado Credit Scoring, e o modelo é conhecido como scorecard de aplicação. O valor limite de uma solicitação de cartão de crédito é determinado usando o scorecard da aplicação, que também ajuda a estimar o nível de risco do solicitante. Esta decisão é tomada com base na prioridade estratégica num determinado momento.

Os clientes devem preencher um formulário, fisicamente ou online, para solicitar um cartão de crédito. Os dados do aplicativo são usados ​​para avaliar a qualidade de crédito do requerente. A decisão é tomada utilizando os dados da aplicação, além do Credit Bureau Score, como o FICO Score nos EUA ou o CIBIL Score na Índia, e outras informações internas sobre os solicitantes. Além disso, os bancos estão rapidamente a ter em conta muitos dados externos para melhorar o calibre dos julgamentos de crédito.

A pontuação do cartão de crédito geralmente é baseada em dados históricos; os modelos preditivos baseados nesses dados estão sujeitos a flutuações econômicas. Portanto, é muito importante sintonizar-se com os dados mais recentes e construir modelos resilientes. Atualmente, com o desenvolvimento de algoritmos de aprendizado de máquina. Métodos mais preditivos, como Boosting, Random Forest e Support Vector Machines, foram introduzidos na pontuação de cartão de crédito.

Construir um modelo de aprendizado de máquina para prever se um candidato é um cliente 'bom' ou 'ruim', diferente de outras tarefas, a definição de 'bom' ou 'ruim' não é dada. Além disso, o problema de desequilíbrio de dados é um grande problema nesta tarefa.

De modo geral, os cartões de pontuação de crédito são baseados em dados históricos. Depois de encontrar grandes flutuações econômicas. Os modelos anteriores podem perder o seu poder preditivo original. O modelo logístico é um método comum para pontuação de crédito. Porque Logística é adequada para tarefas de classificação binária e pode calcular os coeficientes de cada recurso. Para facilitar a compreensão e a operação, o cartão de pontuação multiplicará o coeficiente de regressão logística por um determinado valor (como 100) e o arredondará.

Construir um modelo de aprendizado de máquina para prever se um candidato é um cliente 'bom' ou 'ruim', diferente de outras tarefas, a definição de 'bom' ou 'ruim' não é dada. Além disso, o problema de desequilíbrio de dados é um grande problema nesta tarefa.