

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
ESCOLA DE ENGENHARIA  
CURSO DE ENGENHARIA DE SISTEMAS

THÚLLIO DEBORTOLI MOREIRA ZANETTI

AUTOMAÇÃO DE PROCESSOS DE NEGÓCIO: ESTUDO DE  
TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA  
ANÁLISE DE CRÉDITO

TRABALHO DE CONCLUSÃO DE CURSO II

BELO HORIZONTE  
2018

THÚLLIO DEBORTOLI MOREIRA ZANETTI

AUTOMAÇÃO DE PROCESSOS DE NEGÓCIO: ESTUDO  
DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL PARA  
ANÁLISE DE CRÉDITO

**Trabalho de Conclusão de Curso apresentado como requisito parcial à obtenção do título de Engenheiro de Sistemas, da Escola de Engenharia da Universidade Federal de Minas Gerais.**

**Orientador: Prof. Frederico Gadelha Guimarães**

BELO HORIZONTE, 2018

# Resumo

Este trabalho apresenta a análise comparativa de três diferentes técnicas de Inteligência Artificial para a solução de problema de classificação de *score* de crédito, sendo elas: Árvore de Decisão, Floresta Aleatória e uma Rede Neural Artificial MLP (*Multilayer Perceptron*). A análise de crédito é um processo de negócio muito comum em instituições bancárias e é um tópico muito popular entre os pesquisadores de Inteligência Artificial e *Machine Learning*. Os modelos de classificação obtidos neste trabalho são comparados em relação a acurácia, AUC (*Area Under the ROC Curve*), taxa de Erro do Tipo I e as informações contidas na saída do modelo, tendo em vista que é muito importante para as instituições de crédito a análise da aprovação ou negativa de um possível cliente. A metodologia utilizada para tratamento e transformação dos dados, assim como os passos do treinamento dos modelos escolhidos e os resultados dos testes para cada algoritmo são apresentados.

**Palavras-chave:** Inteligência Artificial, IA, Floresta Aleatória, Árvore de decisão, MLP, RNA, análise de crédito.



# Abstract

The work presented here makes a comparative approach of three different Artificial Intelligence techniques to solve the credit scoring classification problem. The models presented here are Decision Tree, Random Forest and Multilayer Perceptron. Credit analysis is a very popular topic among Artificial Intelligence and Machine Learning researchers. The classification models obtained are compared by their accuracy, AUC (Area Under the ROC Curve), type I error rate and the model's explainability, bearing in mind the importance of credit rejection or approval analysis to credit institutions. The data cleaning methodology, the steps of models training and the results obtained are presented here.

**Keywords:** Artificial Intelligence, AI, Random Forest, Decision Tree, MLP, ANN, credit scoring.



# Lista de abreviaturas e siglas

EUA - Estados Unidos da América

CART - *Classification And Regression Tree* ou, em tradução livre, árvore de classificação e regressão.

RNA - Redes Neurais Artificiais

IA - Inteligência Artificial

VP - Verdadeiro Positivo

VN - Verdadeiro Negativo

FP - Falso Positivo

FN - Falso Negativo





# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
<b>1.1</b>	<b>Problemática</b>	<b>11</b>
1.1.1	Contexto Social	12
<b>1.2</b>	<b>Objetivo</b>	<b>12</b>
1.2.1	Objetivo Geral	12
1.2.2	Objetivos Específicos	13
<b>1.3</b>	<b>Organização</b>	<b>13</b>
<b>2</b>	<b>REFERENCIAL TEÓRICO</b>	<b>15</b>
<b>2.1</b>	<b>Definição</b>	<b>15</b>
2.1.1	Estado da Arte	15
<b>2.2</b>	<b>Desbalanceamento de classes</b>	<b>16</b>
2.2.1	<i>SMOTE : Synthetic Minority Over-sampling Technique</i>	16
<b>2.3</b>	<b>Técnicas de inteligência artificial</b>	<b>17</b>
2.3.1	Árvore de decisão	17
2.3.2	Floresta aleatória	18
2.3.3	Rede neural artificial	20
<b>2.4</b>	<b>Validação cruzada k-fold</b>	<b>21</b>
<b>2.5</b>	<b>Métricas para comparação de resultados</b>	<b>22</b>
2.5.1	Acurácia e Precisão	23
2.5.2	<i>Recall e AUC (Area Under the ROC Curve)</i>	23
2.5.3	Taxa de Erro do Tipo I	24
<b>3</b>	<b>METODOLOGIA</b>	<b>25</b>
<b>3.1</b>	<b>Tratamento e transformação dos dados</b>	<b>25</b>
<b>3.2</b>	<b>Técnicas de inteligência artificial</b>	<b>26</b>
<b>4</b>	<b>RESULTADOS</b>	<b>27</b>
<b>4.1</b>	<b>Árvore de decisão sem <i>oversampling</i></b>	<b>27</b>
<b>4.2</b>	<b>Árvore de decisão com <i>oversampling</i></b>	<b>28</b>
<b>4.3</b>	<b>Floresta aleatória</b>	<b>30</b>
<b>4.4</b>	<b>Rede neural artificial</b>	<b>32</b>
<b>5</b>	<b>CONCLUSÃO</b>	<b>35</b>
5.0.1	Trabalhos futuros	35

<b>REFERÊNCIAS</b> .....	<b>37</b>
--------------------------	-----------

# 1 Introdução

A capacidade de aumento da eficiência de tempo e tomada de decisão dos modelos estatísticos de *score* de crédito no setor financeiro tem atraído muitos pesquisadores nos últimos anos. Estes modelos são dados como indispensáveis em países desenvolvidos e são extremamente necessários nos países em desenvolvimento, para auxílio na análise das políticas de crédito dos bancos. [Abdou et al. 2016]

Dando sequência ao primeiro trabalho realizado [Zanetti 2018], que analisa a importância da automação de processos de negócio e a inserção de inteligência computacional nas diversas atividades das empresas no dia de hoje, este trabalho busca implementar a automação de um processo bastante comum nos bancos, que é o processo de concessão de empréstimo para pessoas físicas. Neste trabalho serão estudadas técnicas de inteligência artificial usadas em análise de crédito para classificar se um cliente se tornará inadimplente ou não, tendo como entrada as informações do cliente e do empréstimo que está sendo solicitado.

## 1.1 Problemática

A análise de risco de crédito bancário exerce uma importante função de classificação entre clientes bons e ruins. Nas últimas décadas tem surgido muitos algoritmos e modelos de classificação para analisar o risco de crédito e predizer se um aplicante ao crédito se tornará inadimplente ou não. Um dos objetivos mais importantes da predição de risco de crédito é a construção do melhor modelo de classificação para uma base de dados específica. [Sang, Nam e Nhan 2016]

Hoje em dia, para qualquer instituição financeira que realiza concessão de crédito é crucial a habilidade de diferenciar um cliente ruim de um cliente bom. A necessidade de modelos confiáveis que classificam os clientes de forma assertiva é muito importante para a sustentabilidade e lucro da empresa. Geralmente, esta predição é realizada através de modelos de classificação, que são treinados a partir de clientes anteriores - tanto clientes adimplentes, quanto inadimplentes - com objetivo de encontrar relações entre características e potencial falha. [Yu, Wang e Lai 2008]

Um ponto que é bastante relevante na análise de crédito é o risco de crédito, que como dito por [Capelletto e Corrar 2008] "é a probabilidade de o tomador dos recursos não pagar ou honrar as obrigações assumidas, tanto no que tange ao principal quanto ao serviço da dívida". Esta preocupação com o melhor gerenciamento do risco de crédito surgiu na década de 90 com mudanças no cenário econômico mundial, como a desregulamentação das

taxas de juros e câmbio, aumento de liquidez e aumento da competição bancária. [Araujo e Carmona 2007]

### 1.1.1 Contexto Social

Uma visão que é muito importante para um Engenheiro de Sistemas é a análise do ambiente no qual o problema está inserido, além do problema técnico em si. No primeiro trabalho [Zanetti 2018], realizou-se a análise socio-econômica do contexto de inserção de inteligência computacional nos processos de negócio das instituições bancárias.

A utilização de inteligência artificial no processo de análise de crédito tem como objetivo analisar as informações de uma pessoa e tomar uma decisão. Este tipo de decisão tem um impacto moral sobre situações do dia-a-dia das pessoas, como se elas conseguirão um empréstimo ou não. Este tipo de problema é muito debatido nos últimos anos em relação ao contexto ético e moral nos quais as máquinas e suas decisões estão inseridas. [Polonski 2018]

Nos últimos anos, além da assertividade, outra característica e ponto de preocupação de muitas empresas é o comportamento de sistemas de *machine learning* como caixa preta: sistemas que recebem uma entrada, produzem uma saída e não dão pistas do porquê. O termo *Explainability*, explicabilidade em tradução livre, é tópico importante para pesquisas na área de IA e que representa várias áreas como viés, transparência, justiça e segurança. [Wierzynski 2018]

## 1.2 Objetivo

A partir dos tópicos citados na seção anterior, são especificados os objetivos deste trabalho.

### 1.2.1 Objetivo Geral

Comparar três diferentes técnicas de inteligência artificial, sendo elas Árvore de Decisão, Floresta Aleatória e RNA, que são utilizadas para classificação de crédito e auxílio à tomada de decisão no segmento financeiro. A comparação será realizada em relação à robustez do modelo, sua assertividade, taxa de erro do tipo I e a quantidade de informações que estão contidas na saída do modelo.

A base de dados utilizada para treinamento e teste do modelo é uma base do *Lending Club*, uma empresa de empréstimos *peer-to-peer* dos EUA, sediada em San Francisco, Califórnia. A base de dados contém informações de empréstimos realizados pela empresa entre 2007 e 2015. A base de dados foi obtida no site *Kaggle* e pode ser acessada em <<https://www.kaggle.com/wendykan/lending-club-loan-data/home>>.

## 1.2.2 Objetivos Específicos

Considerando o desenvolvimento do trabalho e o objetivo geral apresentado, destacam-se os seguintes objetivos específicos:

- Pesquisar sobre técnicas de classificação de crédito
- Buscar uma base de dados confiável e satisfatória para o objeto de estudo em questão
- Trabalhar a base de dados, realizando a limpeza e transformação
- Desenvolver os algoritmos de treinamento e testes das técnicas de IA escolhidas
- Treinar os modelos de cada uma das técnicas
- Realizar testes nos modelos obtidos
- Analisar os resultados obtidos

## 1.3 Organização

O presente trabalho está organizado da seguinte forma:

Capítulo 2: os conceitos relacionados a *score* de crédito, as técnicas utilizadas para tratamento dos dados e os modelos de IA utilizados serão abordados neste capítulo.

Capítulo 3: a metodologia executada para a obtenção dos resultados deste trabalho é apresentada.

Capítulo 4: os resultados dos treinamentos, as análises de dados e os testes realizados serão apresentados neste capítulo.

Capítulo 5: neste capítulo é realizada a conclusão do trabalho levando em consideração os resultados obtidos e os comparativos escolhidos, assim como recomendações para possíveis trabalhos futuros.



## 2 Referencial Teórico

### 2.1 Definição

*Score* de crédito, do inglês *Credit Scoring*, é um conjunto de modelos que permitem que credores, que são pessoas ou instituições que concedem crédito, realizem a análise da concessão de crédito a um cliente. Estes modelos analisam a probabilidade de um devedor se tornar inadimplente ou não e esta técnica é utilizada para decidir quem pode receber um crédito, quanto uma pessoa pode receber de crédito e quanto e como a pessoa terá de pagar pelo crédito. [Thomas, Edelman e Crook 2017]

O *score* de crédito pode ser utilizado para realizar a análise de uma aplicação de crédito, sendo este o objeto de estudo deste trabalho. Neste tipo de sistema são utilizadas informações pessoais e financeiras de quem está aplicando-se ao crédito para cálculo do *score*. Se o *score* estiver acima de um nível seguro, o aplicante tem grande probabilidade de apresentar um bom comportamento de crédito. Já se o *score* estiver abaixo deste limiar, é maior a probabilidade deste aplicante se tornar inadimplente. [Zhao et al. 2015]

#### 2.1.1 Estado da Arte

Como citado por [Abdou et al. 2016], técnicas de regressão logística vem sendo bastante utilizadas para construção de modelos de classificação de crédito, que predizem se um aplicante ao crédito terá um bom comportamento ou não, e são comumente comparadas com métodos não paramétricos, como o CART. Técnicas mais avançadas, e que também são bastante utilizadas, são modelos de redes neurais artificiais.

Nos últimos anos tem crescido o interesse em outros tópicos relacionados à concessão de crédito, como *loss-given default* (LGD) e *exposure at default* (EAD), ou em tradução livre, perda dado a inadimplência e exposição à inadimplência. Apesar deste interesse o tópico de probabilidade de inadimplência e seus modelos de predição, que é o foco deste trabalho, é muito popular e existem muitos trabalhos sobre o mesmo. [Lessmann et al. 2015]

Os resultados para modelos de classificação de crédito variam de acordo com a base de dados e as técnicas utilizadas, como comumente ocorre para trabalhos que envolvem a utilização de algoritmos de inteligência artificial. A assertividade dos modelos podem variar de cerca de 75% (e.g. 76.1% [Bekhet e Eletter 2014]) até mais de 90% (e.g. 92.32% [Abdou et al. 2016]).

## 2.2 Desbalanceamento de classes

Um grande obstáculo dos problemas de classificação e suas base de dados é que em alguns trabalhos não é levado em consideração é o desbalanceamento de classes. O desbalanceamento de classes corresponde quando uma classe é representada por muitos exemplos, enquanto a outra classe é representado apenas por alguns exemplos [Japkowicz 2000]. Japkowicz também cita que este problema é recorrente em vários tipos de problemas e é gargalo para o desempenho dos métodos.

Como será mostrado posteriormente nos resultados deste trabalho, a base de dados escolhida apresenta desbalanceamento entre suas classes, na qual apenas 18.28% dos casos são de clientes ruins (inadimplentes). Pela facilidade de implementação, sua popularidade na literatura e resultados obtidos anteriormente, a técnica utilizada neste trabalho será o SMOTE.

### 2.2.1 SMOTE : *Synthetic Minority Over-sampling Technique*

O algoritmo SMOTE é uma técnica de pré-processamento de dados que realiza o *oversampling* da classe minoritária, isto é, aumenta o número de amostras da classe que têm o menor número de exemplos. O SMOTE tem o objetivo de obter impacto positivo na fronteira de decisão após os treinamentos dos dados, e para isso gera um aumento no número de amostras da classe minoritária por meio da criação de amostras sintéticas. [Marques 2015]

Em [Chawla et al. 2002] são citados os passos para gerar as novas amostras sintéticas. Primeiramente são selecionados os  $k$  vizinhos mais próximos para cada uma das amostras minoritárias e são selecionados  $n$  destes vizinhos para criação das amostras sintéticas. O valor de  $n$  é definido a partir do quanto quer se aumentar a amostra, por exemplo se deseja-se dobrar o numero de amostras, então apenas dois dos  $k$  vizinhos mais próximos de cada amostra original são selecionados para a criação das amostras sintéticas. Então, para cada um dos  $n$  vizinhos de cada uma das amostras são seguidos os seguintes passos:

1. calcula-se a diferença entre a amostra original e seu vizinho
2. multiplica-se este valor da diferença por um número aleatório entre 0 e 1
3. o valor calculado é adicionado a amostra original, obtendo-se a amostra sintética

A Figura 1 mostra graficamente o algoritmo SMOTE atuando para triplicar o número de amostras minoritárias (círculos pretos preenchidos). A amostra  $X$  é selecionada para criação das amostras sintéticas. Os  $k$  vizinhos neste caso são 5 ( $X1$  a  $X5$ ). Destes, apenas dois vizinhos são selecionados, pois a amostra será 200% aumentada. Então acontece



a criação das amostras sintéticas  $Y_1$  e  $Y_2$ . Este processo será repetido para cada uma das amostras originais da classe minoritária e então as classes estarão balanceadas.

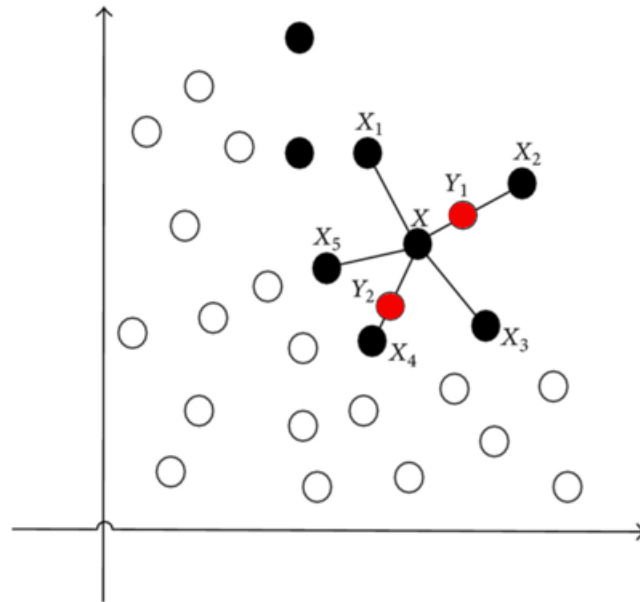


Figura 1 – Modelo Conceitual SMOTE

## 2.3 Técnicas de inteligência artificial

As três técnicas de inteligência artificial que são comparadas neste trabalho são apresentadas a seguir.

### 2.3.1 Árvore de decisão

Árvore de decisão é uma técnica de *machine learning* que consiste em nós de decisão e nós folha. Cada nó de decisão representa um teste  $X$  sobre uma característica da entrada e possui ramificações para cada um destes testes, que representam a separação dos dados de acordo com o resultado do teste. Cada nó folha representa uma classe que é resultado da decisão de cada caso [Stein et al. 2005].

Na Figura 2 pode-se observar um exemplo simplificado de uma estrutura de uma árvore de decisão utilizado na decisão de crédito.

O treinamento de uma árvore de decisão é realizado através de divisão binária recursiva. Consiste em um algoritmo guloso que, em cada passo do algoritmo, são consideradas todas as características em cada ponto de divisão da árvore. Cada uma das possibilidades é testada utilizando uma função de custo, e a divisão que apresenta menor custo para a árvore final é selecionada [Gupta 2017]. Estes testes e decisões são repetidos

até que seja alcançada a condição de parada, que no caso deste trabalho é a profundidade máxima da árvore.

Uma função de custo utilizada em classificadores é o índice de Gini. O cálculo deste índice está presente na Equação 2.1, sendo que  $P_k$  é a fração de classificações corretas para cada uma das classes do problema. O índice Gini mede o quão boa é uma divisão em relação ao quanto as classes estão misturadas nos grupos criados pela divisão [Sanjeevi 2017].

$$Gini = 1 - \sum_{k=0}^{k=1} P_k^2 \quad (2.1)$$

Uma forma de melhorar a performance da árvore é utilizar a técnica de *pruning*. Isto consiste na remoção de ramos da árvore que utilizam características com baixa importância, fazendo com que o poder de predição da árvore aumente, diminuindo o *overfitting* da mesma [Gupta 2017].

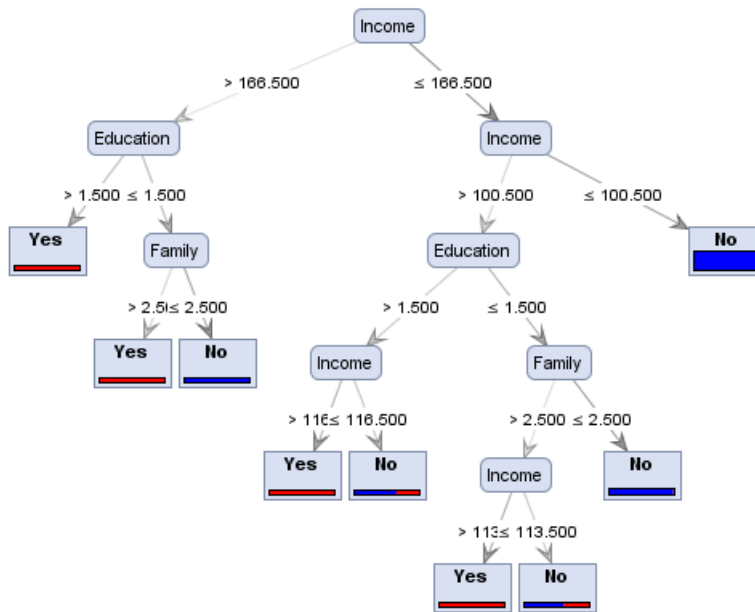


Figura 2 – Modelo conceitual de árvore de decisão

### 2.3.2 Floresta aleatória

Floresta aleatória é um método de *ensemble learning*, em tradução livre aprendizado em conjunto - métodos que combinam o resultados de diversos classificadores gerados. É bastante utilizado atualmente e consiste em um grupo de árvores de decisão, treinadas em amostras *bootstrap* - método de reamostragem que seleciona, aleatoriamente, conjuntos de amostras da base de dados a cada iteração - do conjunto de treinamento e utilizando de

seleção aleatória de característica no processo de geração das árvores. Após a geração de várias árvores, cada uma realiza a classificação para uma dada entrada e a mais popular é escolhida [Brown e Mues 2012].

O treinamento de uma floresta aleatória consiste na criação de várias árvores de decisão, tendo cada uma delas uma profundidade máxima. O número de árvores de decisão utilizadas para o modelo é variado e escolhido o que apresenta melhor acurácia. Isto faz com que o modelo seja mais robusto, porém computacionalmente mais caro.

Na Figura 3 é possível observar o modelo conceitual de uma floresta aleatória e suas árvores de decisão.

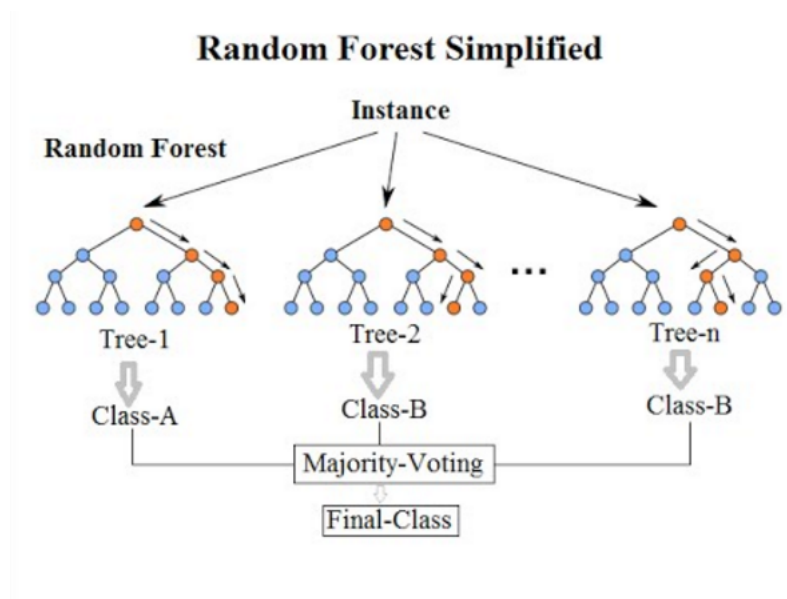


Figura 3 – Modelo conceitual de floresta aleatória

### 2.3.3 Rede neural artificial

Neste trabalho é utilizado uma rede neural do tipo *perceptron* multicamadas (MLP). Esta rede é do tipo *feed-forward* - tipo este no qual as conexões entre as camadas da rede não possui ciclos - e possui 3 camadas. A arquitetura de uma rede MLP genérica pode ser vista na figura 4, que como pode ser observado, possui uma camada de entrada (cada círculo representa uma característica), uma camada escondida (cada círculo representa um neurônio da camada) e uma camada de saída na qual têm-se a probabilidade de cada uma das classes, no caso de problemas de classificação. A rede neural obtida pelo trabalho [Zhao et al. 2015] foi tida como base e apresenta as seguintes características:

- utilização de algoritmo de *backpropagation* (BP) - utilização do erro do modelo para atualização dos pesos e *bias* de cada um dos neurônios - para treinamento, usando o método de gradiente descendente para calcular os novos valores de pesos de acordo com a diferença da saída atual com a saída esperada
- a rede tem uma camada de entrada, uma camada escondida de 9 neurônios e uma camada de saída com 2 neurônios que representa cada uma das classes (créditos bons e ruins)
- a função de ativação dos neurônios é tangente hiperbólica

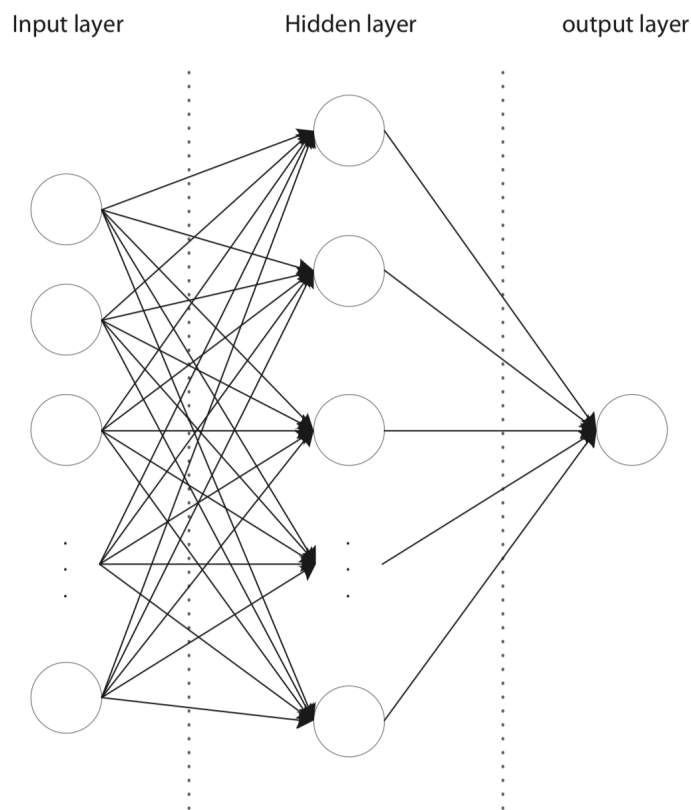


Figura 4 – Modelo conceitual de MLP

## 2.4 Validação cruzada $k$ -fold

Para realizar o treinamento dos modelos propostos, foi utilizada a técnica de validação cruzada  $k$ -fold. Esta técnica visa eliminar o viés do conjunto de treinamento e é bastante utilizada no treinamento de algoritmos de *machine learning* [Brownlee 2018]. Os passos para implementação são:

1. permutar as amostras de forma aleatória
2. dividir a base de dados em  $k$  grupos de tamanhos iguais, ou próximos
3. para cada um dos  $k$  grupos:
  - a) separar o grupo para validação
  - b) treinar o modelo com os grupos restantes
  - c) validar o modelo com o grupo de validação e armazenar o resultado
  - d) descartar o modelo obtido
4. resumir o resultado e obter a performance do modelo a partir da média dos resultados

## 2.5 Métricas para comparação de resultados

Nesta seção serão utilizadas as seguintes siglas para representação das equações:

1. total: número total de predições
2. VP: Verdadeiro Positivo, ou seja, amostras classificadas corretamente como verdadeiras
3. FP: False Positivo, ou seja, amostras classificadas corretamente como falsas
4. VN: Verdadeiro Negativo, ou seja, amostras classificadas erroneamente como verdadeiras
5. FN: False Negativo, ou seja, amostras classificadas erroneamente como falsas

Como citado por [Lessmann et al. 2015], em geral, as performances dos modelos de inteligência artificial são mensurados a partir de três métricas, sendo elas:

1. a capacidade discriminatória - habilidade de separação correta dos dados em classes - do modelo (e.g. AUC)
2. a acurácia do modelo, que mede o quanto o resultado da classificação do modelo representa a realidade
3. A precisão do modelo, que está relacionada a efeitos aleatórios, que é mensurada a partir das taxas de erro do modelo

Sobre as taxas de erros dos modelos, como dito em [Kurum, Yildirak e Weber 2012], existem dois tipos de erros que os modelos de IA estão sujeitos em problemas de classificação, e como no trabalho em questão o objetivo da classificação é prever se o cliente será adimplente (1 para bom pagador e 0 para mal), tem-se:

- Erro Tipo I: Conceder empréstimo para um aplicante que se tornará inadimplente, ou falso positivo
- Erro Tipo II: Recusar o crédito para um aplicante que seria bom pagador, ou falso negativo

Na maioria dos casos, o credor sempre estará mais preocupado com o erro do Tipo I, pois é o erro que tem maior impacto financeiro para a instituição.

Levando os tópicos levantados anteriormente em consideração e os pontos levantados por [Wierzynski 2018] sobre a *Explainability* do modelo, neste trabalho serão considerados os seguintes aspectos com a finalidade de comparação dos modelos.

### 2.5.1 Acurácia e Precisão

Buscando saber a frequência na qual os resultados obtidos pelo classificador estão corretos, a acurácia do modelo é calculada pela Equação 2.2. Outra métrica importante para classificadores, e que frequentemente é confundida com a acurácia, é a precisão do modelo. A precisão busca medir a fração de classes identificadas como verdadeiras realmente eram verdadeiras, e pode ser calculada pela Equação 2.3. [Leal 2017]

$$acc = \frac{VP + VN}{total} \quad (2.2)$$

$$Precision = \frac{VP}{VP + FP} \quad (2.3)$$

### 2.5.2 Recall e AUC (Area Under the ROC Curve)

O *Recall* é a frequência em que o seu classificador encontra os exemplos de uma classe. [Leal 2017] Neste trabalho será utilizado para calcular a frequência que a classe de adimplentes é encontrada pelo classificador e é calculada pela Equação 2.4.

$$Recall = \frac{VP}{VP + FN} \quad (2.4)$$

Outra métrica que é utilizada para observar a capacidade discriminatória dos classificadores é a AUC, que é a probabilidade de um modelo classificar uma amostra aleatória positiva maior que uma amostra aleatória negativa. Esta métrica é calculada como a área da curva ROC do modelo, que representa a performance do modelo plotando dois eixos: VP e FP [Fawcett 2006]. Um exemplo do cálculo do AUC pode ser visto na Figura 5.

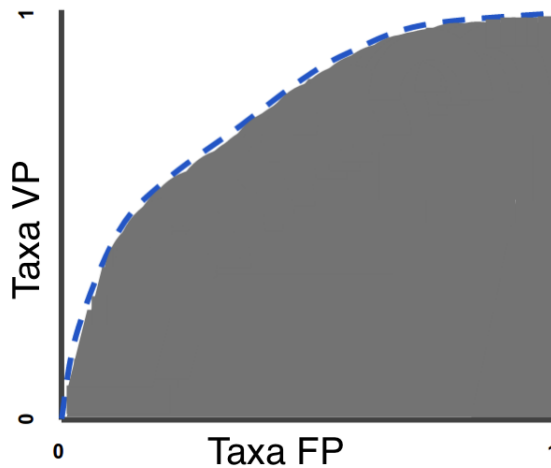


Figura 5 – Cálculo da área sob a curva ROC (AUC)

### 2.5.3 Taxa de Erro do Tipo I

Como já dito anteriormente, para as instituições bancárias é muito prejudicial terem prejuízos e as mesmas, geralmente, tem um limite de taxa de inadimplência que buscam seguir. Por isso a importância de medir-se a taxa de erro do tipo I, que pode ser calculada pela Equação 2.5.

$$errotipoI = \frac{FP}{total} \quad (2.5)$$

A última métrica que será avaliada será a *Explainability* ou ,em tradução livre, explicabilidade do modelo, que é o quão explicativa a saída do modelo é para quem está utilizando-o. Buscando quantificar as informações contidas nas saídas do modelo, visto que é importante para as instituições terem conhecimento do que é produzido pelo modelo, neste trabalho será analisada a saída produzida pelo modelo, observando a rastreabilidade que a mesma oferece em relação à classificação realizada.



## 3 Metodologia

Os passos do desenvolvimento deste trabalho podem ser divididos em duas grandes partes: primeiramente o tratamento e transformação da base de dados, pois o mesmo possui mais informações que o necessário e informações que devem ser tratadas antes dos algoritmos serem treinados. Após isto os algoritmos de inteligência artificial são treinados e os modelos e resultados obtidos. As seções deste capítulo apresentam os passos e desafios encontrados neste desenvolvimento.

### 3.1 Tratamento e transformação dos dados

Como já citado anteriormente a base de dados obtida da *Lending Club* é uma base histórica com todas as operações de empréstimo realizadas através da empresa.

Primeiramente foram filtradas as amostras que apresentam *status* finais, isto é que ou já foram integralmente pagos ou que a operação já reportada como prejuízo, pois como a análise deste trabalho levará em consideração se o aplicante entrará em inadimplência ou não, os status nos quais o devedor ainda se encontra em pagamento das parcelas não tem informação necessária para saber se entrará como prejuízo para a instituição ou não.

Depois disso são filtradas as características nas quais mais de 50% das amostras não apresentam informação e as características que são posteriores a concessão do empréstimo, tendo em vista que não têm informação necessária para o problema e que a predição ocorrerá na fase de aplicação ao crédito.

Posteriormente as características categóricas (textuais) são transformadas em numéricas para realizar o treinamento dos modelos. Algumas variáveis ainda sofrem alteração para apresentarem valores mais informativo ao problema e são filtradas em relação a correlação com a saída.

Após este processo, são obtidas 15 características que serão utilizadas para treinamento dos modelos. Desta maneira, os dados são utilizados para realizar o treinamento da árvore de decisão e da floresta aleatória.

Para a RNA os dados tiveram de sofrer mais uma alteração. Baseado em [Anis 2015] e [Ng] foi realizado a normalização *Z-score*, pois os dados não apresentam escala e distribuição uniforme, o que é muito importante para RNA para que a convergência do algoritmo ocorra de maneira mais rápida. A normalização escolhida busca dispor os dados de forma que apresentem média igual a 0 e normalizar a variância de cada característica. A normalização é realizada através da equação 3.1, sendo  $x$  os dados  $\mu$  a média das amostras

e  $\sigma^2$  a variância.

$$y = \frac{x - \mu}{\sigma^2} \quad (3.1)$$

## 3.2 Técnicas de inteligência artificial

Para treinamento dos modelos de inteligência artificial, Primeiramente 10% das amostras foram separadas para realizar os testes para comparação dos modelos. Após isto, as amostras foram separadas em 10 grupos para ser executado o método de validação cruzada *k-fold*. Com os dados da maneira que foram obtidos, como será mostrado nos resultados, foi descoberto que a base de dados apresentava o problema de classes desbalanceadas, então no conjunto de treinamento e validação foi-se aplicada a técnica *SMOTE* para realizar o *oversampling* das amostras do conjunto minoritário. Assim a base de dados foi separada para realizar o treinamento dos modelos.

Para o treinamento dos modelos de árvore de decisão e floresta aleatória foram variados alguns hiperparâmetros para encontrar os que obtiam melhor performance. Para a árvore de decisão foi variado a profundidade da árvore e para a floresta aleatória foram variados a profundidade de cada árvore e o número de classificadores (árvores de decisão) utilizados. Para cada técnica foram escolhidos os classificadores com hiperparâmetros que obtiveram melhor performance.

Para a RNA, como dito anteriormente, antes de realizar o treinamento da rede, as amostras foram normalizadas. Baseando-se nos passos e parâmetros para construção da rede de [Zhao et al. 2015], realizou-se o treinamento e obtenção dos pesos e *bias* da rede.

Obtendo-se os classificadores, então as amostras separadas para teste foram avaliadas em cada um deles nas métricas estabelecidas anteriormente para avaliação. Buscando-se eliminar o viés de conjunto de amostras e obter resultados mais representativos, este processo foi repetido 30 vezes para cada um dos algoritmos e a avaliou-se performance média de cada um dos modelos.

Nos próximos capítulos são apresentados os resultados obtidos e a comparação dos mesmos.

## 4 Resultados

Para cada um dos modelos buscou-se apresentar os resultados em relação a acurácia média, a média do percentual de erros dos tipos I e II, a média do *AUC score* e a representação da quantidade de informações das saídas dos modelos. Para complementar os resultados, para alguns modelos foi-se obtida a importância de cada uma das características para a construção do modelo. Os resultados são analisados a seguir.

### 4.1 Árvore de decisão sem *oversampling*

Primeiramente, obteve-se um modelo de de árvore de decisão sem realizar-se o *oversampling* das amostras previamente. Com isto obteve-se um modelo com resultados bastante expressivos, com mais de 90% de acurácia, como mostra a Figura 6.

Porém, analisando a estrutura da árvore, mostrada na Figura 7, e o resultado de 0% de erros do tipo II é possível identificar a caracterização de um problema de desbalanceamento de classe, no qual apenas uma característica é utilizada para a classificação. Com isto a classe majoritária é selecionada na maior parte dos casos, obtendo-se um erro do tipo II de 0% e uma acurácia elevada.

```
Mean Type I error (FP): 0.08780  
Mean Type II error (FN): 0.00000  
Mean Accuracy: 0.91220  
Mean ROC_AUC_Score: 0.83788
```

Figura 6 – Resultados da árvore de decisão sem *oversampling*

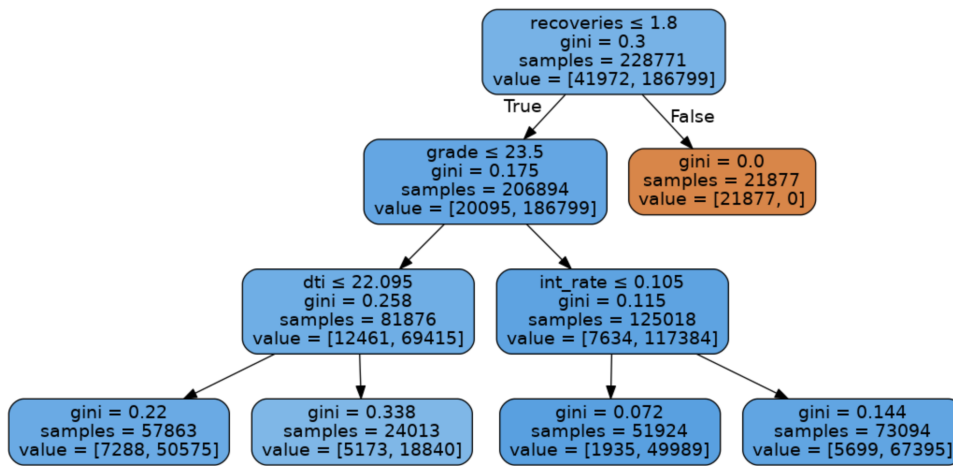


Figura 7 – Estrutura da árvore de decisão sem *oversampling*

## 4.2 Árvore de decisão com *oversampling*

Após a realização do *oversampling* das amostras da classe minoritária a nova estrutura da árvore de decisão apresentou profundidade média de 18.3, apresentando assim uma estrutura mais coerente e mais robusta, apesar de apresentar resultados piores, como esperado.

A Figura 8 apresenta os resultados do modelo e a Figura 9 apresenta alguns dos nós de um dos melhores classificadores obtidos para a técnica e representa uma possível representação da saída do modelo, na qual pode-se observar qual o caminho que resulta na aprovação ou reprovação de um determinado aplicante.

Mean Type I error (FP): 0.23303  
 Mean Type II error (FN): 0.00987  
 Mean Accuracy: 0.75710  
 Mean ROC\_AUC\_Score: 0.83447

Figura 8 – Resultados da árvore de decisão com *oversampling*

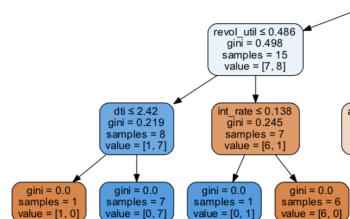


Figura 9 – Estrutura de alguns nós da árvore de decisão com *oversampling*

Uma informação importante do modelo é a importância de cada uma das 15 características para o modelo. Como pode ser visto na Figura 10, uma das características apresenta uma importância dominante em relação as outras, sendo que algumas nem sequer são utilizadas pelo modelo e poderiam ser descartadas.

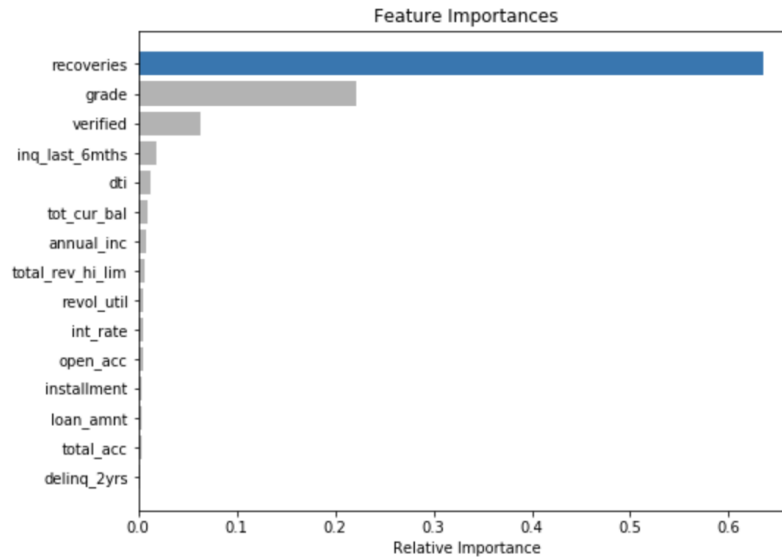


Figura 10 – Importância normalizada de cada uma das características

Na Figura 11 são observados os valores médios da acurácia para os conjuntos de validação utilizados na validação cruzada para cada um dos hiperparâmetros comparados. É possível observar que a acurácia eleva-se e mantém-se praticamente constante em cada dois hiperparâmetros, sendo maior em profundidades pares, apresentando valores acima de 90% de acurácia para o conjunto de validação em alguns casos.

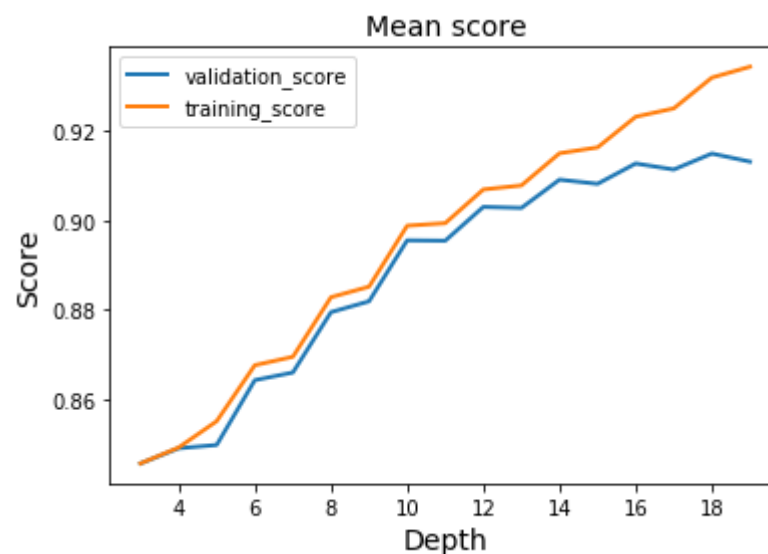


Figura 11 – Acurácia média para cada um dos hiperparâmetros testados

### 4.3 Floresta aleatória

Para o modelo de Floresta Aleatória os hiperparâmetros médios encontrados nas 30 execuções foram: profundidade máxima de cada uma das árvores de decisão igual a 15 e número médio do número de classificadores utilizado na Floresta aleatória igual a 44. A acurácia média para cada um dos conjuntos de hiperparâmetros pode ser visto na Figura 12.

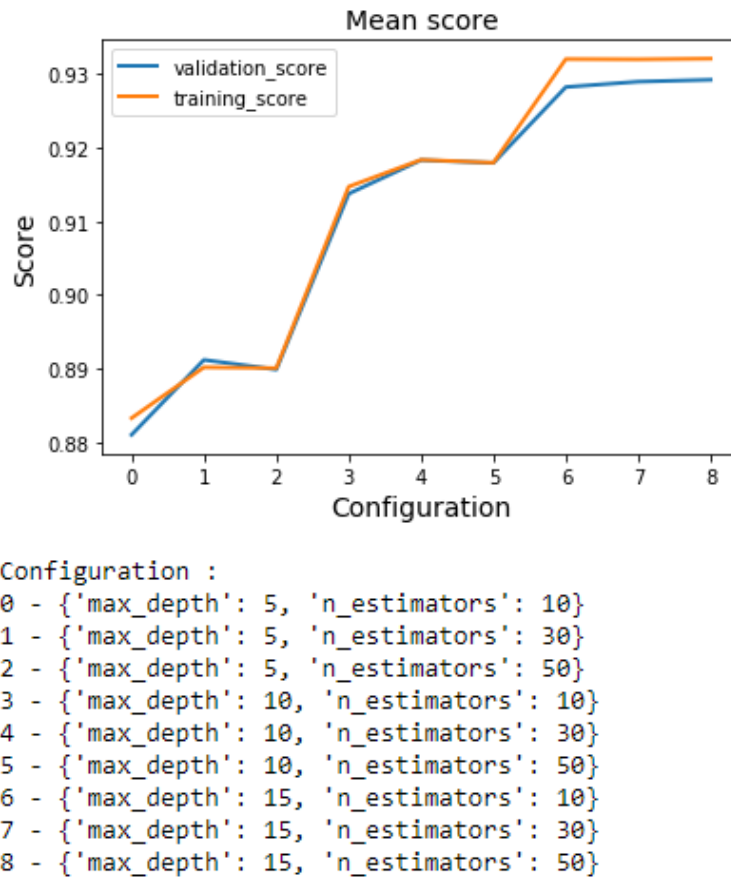


Figura 12 – Acurácia média para cada um dos hiperparâmetros testados

Como mostrado pelas Figuras 13 e 14, os resultados apresentados pela floresta aleatória e a estrutura da saída são muito semelhantes aos resultados da árvore de decisão, tendo pequena vantagem em relação à acurácia e a capacidade discriminatória (*AUC score*).

```

Mean Type I error (FP): 0.23889
Mean Type II error (FN): 0.00013
Mean Accuracy: 0.76098
Mean ROC_AUC_Score: 0.86164

```

Figura 13 – Resultados da floresta aleatória

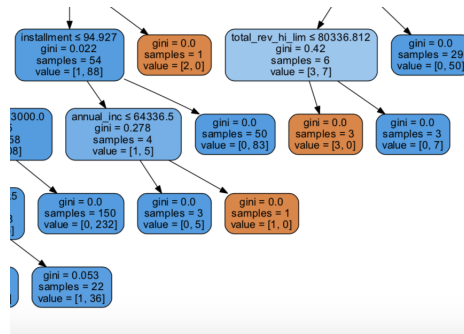


Figura 14 – Estrutura de alguns nós de uma das árvores de decisão da floresta aleatória

Uma característica importante da floresta aleatória é que todas as características foram consideradas pelo classificador, apesar de algumas terem uma importância muito inferior as outras, como mostrado pela Figura 15. Esta característica deve-se ao fato da utilização de vários classificadores.

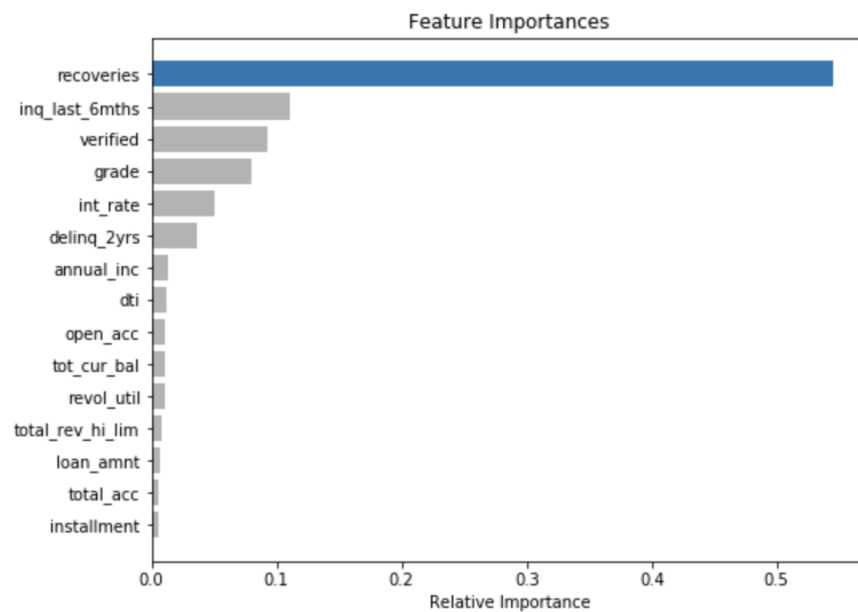


Figura 15 – Importância normalizada de cada uma das características

## 4.4 Rede neural artificial

Conforme citado anteriormente, a saída dos modelos de RNA são como caixas pretas, pois os valores que levam uma entrada ser aprovada ou reprovada não tem muita informação para quem concede o crédito, sendo definida pelos valores dos pesos e *bias* de cada um dos neurônios e suas funções de ativação, semelhante ao que é mostrado na Figura 16. Os resultados observados na Figura 17 mostram que apesar da RNA apresentar uma acurácia média menor que os outros classificadores, apresenta uma taxa de erro do tipo I significativamente menor (cerca de 5% menor que os outros modelos).

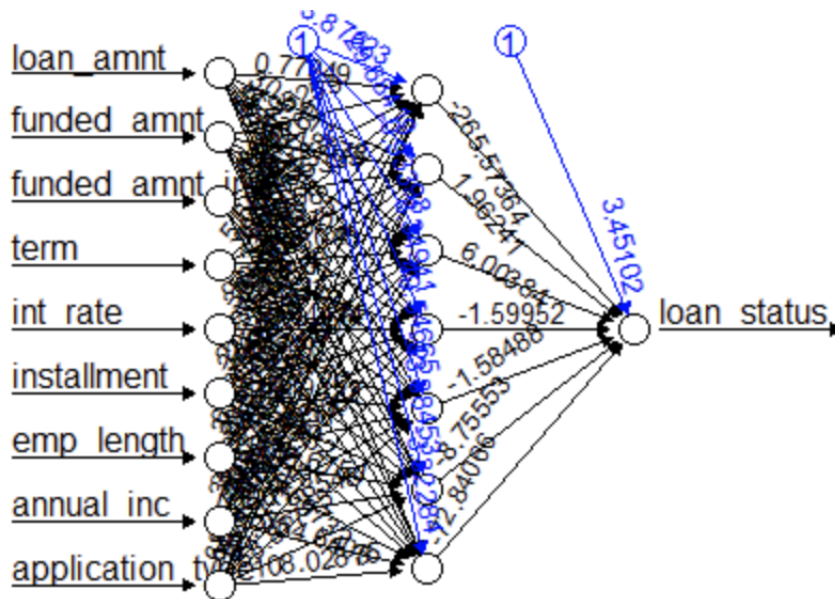


Figura 16 – Exemplo de saída de uma RNA

```

Mean Type I error (FP): 0.17934
Mean Type II error (FN): 0.11204
Mean Accuracy: 0.70862
Mean ROC_AUC_Score: 0.78674

```

Figura 17 – Resultados da RNA

A Figura 18 mostra que não houve enviesamento no treinamento do modelo, com a curva de aprendizado do conjunto de validação do *k* – *fold* sendo semelhante ao conjunto de treinamento.



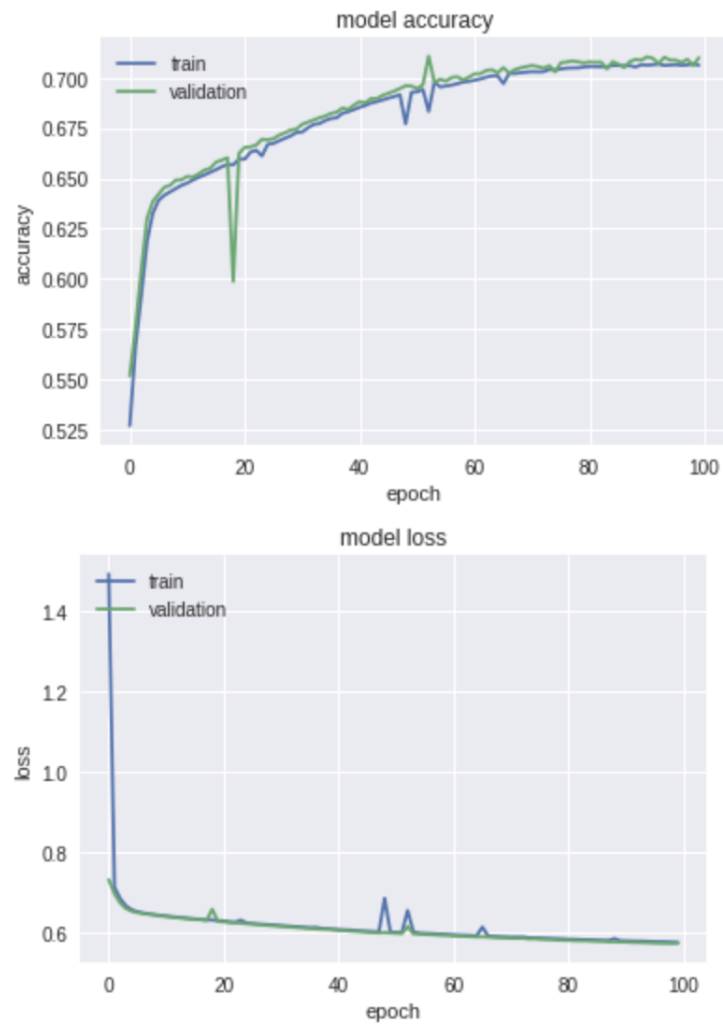


Figura 18 – Curvas de aprendizado dos conjuntos de treinamento e validação

A Figura 19 apresenta os resultados dos modelos em relação às métricas de *Recall* e Precisão dos modelos. É possível observar que os modelos de Árvore de decisão e Floresta Aleatória apresentam maior *Recall*, o que representa que elas geralmente classificam os clientes adimplentes corretamente. Porém, quando analisada a precisão dos modelos é possível perceber que a RNA apresenta um resultado melhor.

Esta característica é comum como citado em [Leal 2017] "o quanto mais rigorosos somos para acertar (melhorar a precisão) menos estamos dispostos a errar (aumentar o recall)".

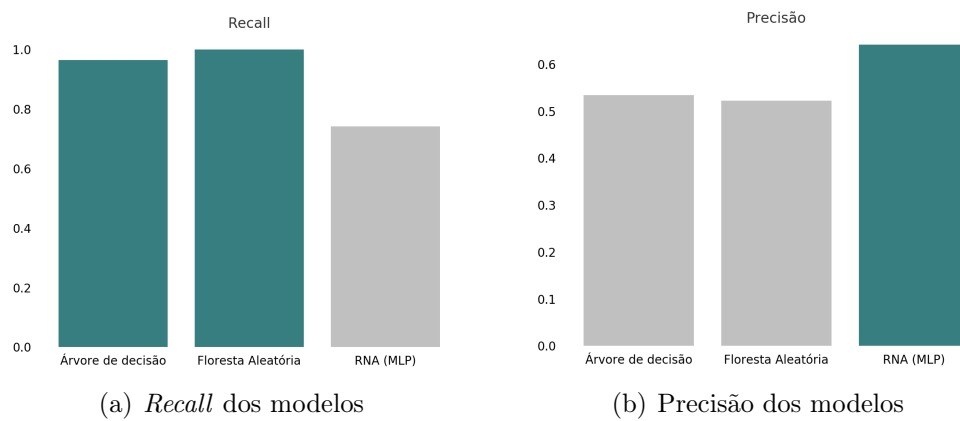
(a) *Recall* dos modelos(b) *Precisão* dos modelos

Figura 19 – Comparativo de métricas dos classificadores

## 5 Conclusão

Com os resultados obtidos pelos três modelos apresentados no capítulo anterior, podemos classificá-los em dois grupos. O primeiro é o da árvore de decisão e floresta aleatória, os quais apresentam uma saída com mais informações que o modelo de RNA e também apresentam uma acurácia e capacidade discriminatória semelhantes - sendo que a floresta aleatória obteve resultados um pouco melhores - e mais expressivos que o outro classificador. Já o modelo de RNA apresentou resultados muito mais expressivo em relação ao erro do tipo I, com o qual é medido a possível taxa de inadimplência do credor, sendo este um importante índice para as instituições financeiras.

Existem dois tipos de métricas, sendo elas métricas *optimizing* e *satisficing*. Métricas *satisficing* são aquelas que tem um limiar e devem ser correspondidas pelo modelo pra que o mesmo seja considerado satisfatório (e.g. o erro do tipo I do modelo deve ser menor ou igual a 0.2). Já métricas *optimizing* são utilizadas para selecionar o melhor modelo dentro aqueles que são considerados satisfatórios [Ng]. Com isto, a tomada de decisão de qual modelo será utilizado fica a cargo da instituição, tendo em vista que o erro do tipo I pode ser uma métrica *satisficing* para a organização.

Na literatura, geralmente, os algoritmos de RNA apresentam resultados mais expressivos que os outros classificadores ([Abdou et al. 2016], [Lessmann et al. 2015]), porém este trabalho não obteve o mesmo resultado. Isto pode ser explicado devido ao fato de que em problemas de classes desbalanceadas a técnica SMOTE não apresentar melhoria significativa nos resultados de redes MLP [Aurelio 2017] e como citado por [Silva 2012], o tamanho da base de dados também influencia nos resultados das florestas aleatórias, fazendo com que as mesmas apresentem alta complexidade, e este foi o caso deste trabalho, no qual após a realização do *oversampling* das amostras, a base de dados ficou com quase 400 mil exemplos.

### 5.0.1 Trabalhos futuros

As propostas de continuidade em trabalhos futuros podem ser divididas em três grupos distintos: maior tratamento e transformação da base de dados, utilização de outra técnica para lidar com o problema de classe desbalanceada e a utilização de diferentes modelos de inteligência artificial.

A primeira trata-se da utilização de diferentes características da base de dados inicial e/ou a criação de novos parâmetros a partir dos existentes, buscando assim obter características que tenham maior correlação com a saída.

A segunda proposta é a de utilização de diferentes técnicas para tratamento de

classes desbalanceadas, tendo em vista que o SMOTE não é a melhor técnica para redes MLP e as mesmas podem apresentar resultados mais expressivos. Outras técnicas que poderiam ser utilizadas como manipulação algorítmica, como apresentado em [Aurelio 2017], buscando melhores resultados com a RNA.

A última trata-se da utilização de diferentes técnicas de IA para a classificação, podendo ser técnicas não utilizadas neste trabalho (e.g. SVM), utilização de diferentes arquiteturas de RNA além da MLP, que é uma das redes mais básicas, ou a utilização de um *ensemble*, que é uma estratégia de combinação de diferentes modelos com objetivo de obter um classificador mais robusto.

# Referências

ABDOU, H. A. et al. Predicting creditworthiness in retail banking with limited scoring data. *Knowledge-Based Systems*, v. 103, p. 89–103, April 2016. Citado 3 vezes nas páginas 11, 15 e 35.

ANIS, A. *A Dummies Guide to Data Normalization*. 2015. Accessed: 2018-10-05. Disponível em: <<https://medium.com/@lahorekid/a-dummies-guide-to-data-normalization-for-neural-nets-ff1998116e75>>. Citado na página 25.

ARAUJO, E. A.; CARMONA, C. U. M. Desenvolvimento de modelos credit scoring com abordagem de regressão logística para a gestão da inadimplência de uma instituição de microcrédito. *Contabilidade Vista & Revista*, v. 18, p. 107–131, Julho/Setembro 2007. Citado na página 12.

AURELIO, Y. S. *Estudo de funções de custo para redes neurais com dados desbalanceados*. Dissertação (Mestrado) — UFMG - Universidade Federal de Minas Gerais, 2017. Citado 2 vezes nas páginas 35 e 36.

BEKHET, H. A.; ELETTER, S. F. K. Credit risk assessment model for jordanian commercial banks: Neural scoring approach. *Review of Development Finance*, v. 4, p. 20–28, 2014. Citado na página 15.

BROWN, I.; MUES, C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications*, v. 39, p. 3446–3453, 2012. Citado na página 19.

BROWNLEE, J. *A Gentle Introduction to k-fold Cross-Validation*. 2018. Accessed: 2018-10-05. Disponível em: <<https://machinelearningmastery.com/k-fold-cross-validation/>>. Citado na página 21.

CAPELLETTO, L. R.; CORRAR, L. J. Índices de risco sistêmico para o setor bancário. *Revista Contabilidade & Finanças*, v. 19, p. 6–18, Maio/Agosto 2008. Citado na página 11.

CHAWLA, N. V. et al. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 2002. Citado na página 16.

FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, p. 861–874, 2006. Citado na página 23.

GUPTA, P. *Decision Trees in Machine Learning*. 2017. Accessed: 2018-11-05. Disponível em: <<https://towardsdatascience.com/decision-trees-in-machine-learning-641b9c4e8052>>. Citado 2 vezes nas páginas 17 e 18.

JAPKOWICZ, N. The class imbalance problem: Significance and strategies. *Proceedings of the 2000 International Conference on Artificial Intelligence (IC-AI'2000): Special Track on Inductive Learning*, Las Vegas, Nevada, 2000. Citado na página 16.

KURUM, E.; YILDIRAK, K.; WEBER, G. A classification problem of credit risk rating investigated and solved by optimization of roc curve. *Centr. Eur. J. Oper. Res.*, v. 20, p. 529–557, 2012. Citado na página 22.

LEAL, R. dos S. *Métricas Comuns em Machine Learning: como analisar a qualidade de chat bots inteligentes*. 2017. Accessed: 2018-11-05. Disponível em: <<https://medium.com/as-m%C3%A1quinas-que-pensam/m%C3%A9tricas-comuns-em-machine-learning-como-analisar-a-qualidade-de-chat-bots-inteligentes-m%C3%A9tricas-1ba580d7cc96>>. Citado 2 vezes nas páginas 23 e 33.

LESSMANN, S. et al. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, v. 247, p. 124–136, 2015. Citado 3 vezes nas páginas 15, 22 e 35.

MARQUES, Y. B. *Miracle: Aprendizagem de máquina utilizando SMOTE e Random Forest para prover aumento da seletividade na predição ab initio de pre-miRNAs*. Dissertação (Mestrado) — UFV - Universidade Federal de Viçosa, 2015. Citado na página 16.

NG, A. Y. *Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization*. Accessed: 2018-10-05. Disponível em: <<https://www.youtube.com/watch?v=Ulp2CMI0748>>. Citado na página 25.

NG, A. Y. *Structuring Machine Learning Projects*. Accessed: 2018-10-05. Disponível em: <<https://www.coursera.org/learn/machine-learning-projects/lecture/uNWnZ/satisficing-and-optimizing-metric>>. Citado na página 35.

POLONSKI, V. *The Hard Problem of AI Ethics - Three Guidelines for Building Morality Into Machines*. 2018. Accessed: 2018-11-29. Disponível em: <<https://www.oecd-forum.org/users/80891-dr-vyacheslav-polonski/posts/30743-the-hard-problem-of-ai-ethics-three-guidelines-for-building-morality-into-machines>>. Citado na página 12.

SANG, H. V.; NAM, N. H.; NHAN, N. D. A novel credit scoring prediction model based on feature selection approach and parallel random forest. *Indian Journal of Science and Technology*, v. 9(20), May 2016. Citado na página 11.

SANJEEVI, M. *Chapter 4: Decision Trees Algorithms*. 2017. Accessed: 2018-11-05. Disponível em: <<https://medium.com/deep-math-machine-learning-ai/chapter-4-decision-trees-algorithms-b93975f7a1f1>>. Citado na página 18.

SILVA, J. C. da. *Aprendendo em uma Floresta Aleatória*. 2012. Accessed: 2018-11-05. Disponível em: <<https://medium.com/machina-sapiens/o-algoritmo-da-floresta-aleatoria-3545f6babdf8>>. Citado na página 35.

STEIN, G. et al. Decision tree classifier for network intrusion detection with ga-based feature selection. *Proceedings of the 43rd ACM Southeast Conference*, Kennesaw, GA, 2005. Citado na página 17.

THOMAS, L. C.; EDELMAN, D. B.; CROOK, J. N. *Credit scoring and its applications*. 2. ed. Philadelphia, PA: [s.n.], 2017. Citado na página 15.

WIERZYNSKI, C. *The Challenges and opportunities of explainable AI*. 2018. Accessed: 2018-10-05. Disponível em: <<https://ai.intel.com/the-challenges-and-opportunities-of-explainable-ai/>>. Citado 2 vezes nas páginas 12 e 22.

YU, L.; WANG, S.; LAI, K. K. Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications*, v. 34, p. 1434–1444, 2008. Citado na página 11.

ZANETTI, T. D. M. Os impactos da automação de processos de negócio nos funcionários de uma instituição bancária. *Universidade Federal de Minas Gerais - UFMG*, 2018. Citado 2 vezes nas páginas 11 e 12.

ZHAO, Z. et al. Investigation and improvement of multi-layer perception neural networks for credit scoring. *Expert Systems with Applications*, v. 42, p. 3508–3516, May 2015. Citado 3 vezes nas páginas 15, 20 e 26.