

## Desenvolvimento de Sistema de Inteligência Artificial (IA) Baseado em Reconhecimento de Padrões Para Análise de Rotas Veiculares

## Development of an Artificial Intelligence System (AI) Based on Patterns Recognition for the Analysis of Vehicular Routes

Leonardo Luís Röpke<sup>[1,A]</sup>, Manuel Osório Binelo<sup>[1,B]</sup>

<sup>[1]</sup> UNIJUÍ – Universidade Regional do Noroeste do Estado do Rio Grande do Sul, Ijuí, RS, Brazil.

<sup>[A]</sup> leonardoropke@gmail.com, <sup>[B]</sup> manuel.binelo@unijui.edu.br

**Abstract** This work presents the study and development of an Artificial Intelligence system, with focus on K-means algorithms and Artificial Neural Networks, to assist fleet managers in the identification of routes and route deviations. The developed tool has the objective of modernizing the process of identification of routes and deviations of routes. The results show that the Artificial Neural Networks obtained a 100% accuracy rate in the identification of routes, and in the identification of route deviations the RNAs were able to identify 61% of the routes presented. Therefore, RNAs are an excellent technique to be applied to the identification of routes and deviations of routes. The K-means algorithm presented good results when applied in the discovery of similar routes, thus becoming an important tool applied to the work of monitoring vehicles routes.

**Resumo** Este trabalho apresenta o estudo e desenvolvimento de um sistema de Inteligência Artificial, com foco nos algoritmos K-means e Redes Neurais Artificiais, para auxiliar gestores de frotas na identificação de rotas e desvios de rotas. A ferramenta desenvolvida tem como objetivo modernizar o processo de identificação de rotas e desvios de rotas. Os resultados mostram que as Redes Neurais Artificiais obtiveram uma taxa de 100% de acerto na identificação de rotas, já na identificação de desvios de rotas as RNAs foram capazes de identificar 61% das rotas apresentadas. Portanto, as RNAs demonstraram ser uma excelente técnica a ser aplicada pra a identificação de rotas e desvios de rotas. O algoritmo K-means apresentou bons resultados ao ser aplicado na descoberta de rotas semelhantes, tornando-se assim uma ótima ferramenta aplicada ao trabalho de monitoramento das rotas percorridas pelos veículos.

**Keywords:** Artificial intelligence, Artificial neural networks, K-means, Pattern Recognition, Routes.

**Palavras-chave:** Inteligência Artificial, K-means, Reconhecimento de Padrões, Redes Neurais Artificiais, Rotas.

## 1 Introdução

A Inteligência Artificial (IA) é um ramo da ciência da computação que se propõe a desenvolver formas de simulação da capacidade humana. Nela são elencadas características como o raciocínio, reconhecimento de padrões, tomada de decisões e solução de problemas. A IA é apontada como um campo de conhecimento que pode apoiar a tomada de decisões de um modo mais simples e mais preciso do que outros métodos, tais como a modelagem e a gestão por indicadores [1].

A utilização da IA associada a técnicas de Data Mining apresenta relacionamentos existentes, que por si só dificilmente seriam identificados em análises simplificadas. Estas técnicas podem ser aplicadas em diversas áreas, como é o caso da logística, onde existem sistemas criados exclusivamente para esta cadeia visando auxiliar no

controle de rotas e de tráfego. Estes sistemas se baseiam em um processo integrado, unindo software e hardware, propiciando uma utilização simples e fácil [2].

Ferramentas assim surgem como uma alternativa de apoio aos métodos tradicionais de tomada de decisão da área de logística. A construção de uma metodologia de apoio à decisão através do uso de técnicas de IA com o intuito de facilitar o controle das atividades logísticas é vantajosa sob o ponto de vista do gerenciamento e também da verificação da aplicabilidade dessas técnicas em relação ao problema proposto [3].

Dois métodos de IA bastante utilizados, o reconhecimento de padrões e a clusterização de dados apresentam formas de solução para os mais variados problemas. Dentro do reconhecimento de padrões as Redes Neurais Artificiais (RNAs) se destacam, sendo apontadas por especialistas de diversas áreas como uma alternativa eficiente aplicada na área de transportes [4]. Já dentro dos algoritmos de clusterização um dos mais conhecidos é o K-means, seu objetivo é encontrar a melhor divisão de uma quantidade P de dados em K grupos, de forma que os dados mais semelhantes estejam agrupados [5].

Atualmente na área de logística pode-se considerar o processo de monitoramento e avaliação de rotas bem exaustivo, contando que o gestor da frota deve monitorar seus veículos um a um a fim de analisar a rota percorrida. Em alguns sistemas de monitoramento são emitidos alertas conforme o veículo percorre determinada rota, mas o mesmo não se aplica ao histórico de trajetos percorridos pelos veículos, que neste caso o usuário deve buscar o histórico e visualizar um a um os trajetos percorridos.

Dentro deste contexto mostra-se de grande valor a análise e comparação de técnicas de IA, além do desenvolvimento de um sistema de IA que avalie de forma autônoma ou semiautônoma as rotas percorridas por veículos de forma semelhante a que um gestor faria.

Este artigo está organizado da seguinte forma. Na seção 2, são explanadas as técnicas de IA abordadas, além da definição de rotas e de rastreamento veicular. A seção 3 aborda os materiais e métodos utilizados para o desenvolvimento do trabalho e obtenção dos resultados, que são apresentados na seção 4. Na seção 5 é apresentada a conclusão do trabalho, com uma síntese sobre os resultados obtidos.

## 1.1 Inteligência Artificial

A Inteligência Artificial (IA) é composta de várias áreas de pesquisa, das quais as mais comuns são: busca (de soluções); sistemas especialistas; processamento de linguagem natural; reconhecimento de padrão; robótica; aprendizado de máquina e lógica [6].

Um sistema de IA não é capaz somente de armazenar e manipular dados, mas também da aquisição, representação, e manipulação do conhecimento. Esta manipulação inclui a capacidade de deduzir ou inferir novos conhecimentos - novas relações sobre fatos e conceitos - a partir do conhecimento existente e utilizar métodos de representação e manipulação para resolver problemas complexos [7].

Os motores de inferência podem ser desenvolvidos ou reutilizados a partir de outros projetos de IA. São eles que determinam os itens de conhecimento a serem acessados, as deduções a serem feitas e a ordem dos passos a serem utilizados. A partir da adoção de um motor de inferência, o desenvolvimento de uma IA é reduzido a obter e codificar regras e fatos necessários para um determinado domínio do problema. Portanto, o projetista de uma IA deve se preocupar principalmente com a aquisição, representação e manipulação do conhecimento. A Figura 1 mostra a inter-relação entre os componentes de um sistema clássico de IA.

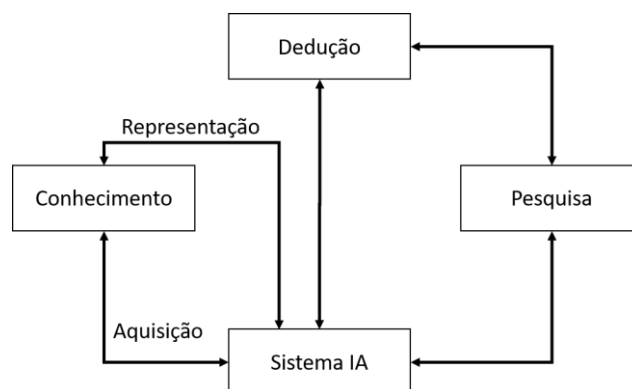


Figura 1. Visão conceitual sobre os sistemas de IA. Fonte: [28].

## 1.2 Redes Neurais Artificiais

As Redes Neurais Artificiais ou RNAs são consideradas uma classe de modelos matemáticos para resolução de problemas de classificação e reconhecimento de padrões, consistindo numa teoria para o estudo de fenômenos complexos [8]. Todas as RNAs são compostas por vários elementos de processamento simples, conhecidos como neurônios [9], sendo que uma grande RNA pode ter centenas ou milhares de unidade de processamento, já o cérebro de um mamífero pode ter muitos bilhões de neurônios [10]. A construção das RNAs tem inspiração nos neurônios biológicos e nos sistemas nervosos. Entretanto, é importante compreender que atualmente as RNAs estão distantes das Redes Neurais Naturais e frequentemente as semelhanças são mínimas [11].

As RNAs consistem em um método de IA para a solução de problemas, através de um sistema que simula o cérebro humano, inclusive em seu comportamento, ou seja, aprendendo, errando e fazendo descobertas. A vantagem das RNAs consiste no fato de resolverem problemas complexos que envolvem o reconhecimento de padrões [12].

A maioria dos modelos de RNAs possui alguma regra de treinamento, através da qual os pesos das conexões são ajustados de acordo com os padrões apresentados. As RNAs aprendem através de exemplos e possuem um grande potencial de paralelismo sendo que uma Rede Neural é especificada, principalmente pela sua topologia, pelas características dos nós, pelas regras de treinamento e parâmetros de treinamento [10]. São muitos os benefícios sobre o uso de RNAs na computação, conforme [13] pode-se citar:

- Habilidade de aprender com exemplos: Os computadores neurais têm a capacidade de aprender com a experiência, objetivando aperfeiçoar seu desempenho e se adaptar aos cenários novos e dinâmicos;
- Robustez: RNAs têm habilidade em tratar com ruídos. Elas são tolerantes a falhas e podem apresentar degradação gradual, ou seja, apesar de existir alguma falha no sistema, elas continuam oferecendo respostas adequadas por um período significativo de tempo;
- Velocidade de processamento: Como as RNAs consistem de grande número de unidades de processamento operando de forma paralela, elas podem operar em velocidades consideráveis em relação aos métodos computacionais comuns.

O neurônio artificial é a unidade fundamental de processamento de uma RNA, o qual recebe uma ou mais entradas, transformando-as em saídas. Os neurônios artificiais são geralmente conectados por canais de comunicação que estão associados a determinados pesos, eles executam operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. O comportamento “inteligente” de uma Rede Neural Artificial vem das interações entre os neurônios artificiais [10].

A operação de um neurônio artificial baseado no modelo proposto por McCulloch e Pitts (1943), pode ser definida da seguinte maneira:

- Valores são apresentados à entrada;
- Cada valor é multiplicado por um peso sináptico, que indica a sua influência na saída;
- É realizada a soma ponderada dos sinais que produz um nível de atividade;
- Na função de ativação, se o nível de atividade atingir o limiar (threshold) a unidade produz uma saída igual a 1; caso contrário, igual a 0.

O modelo de McCulloch e Pitts pode ser visto na Figura 2.

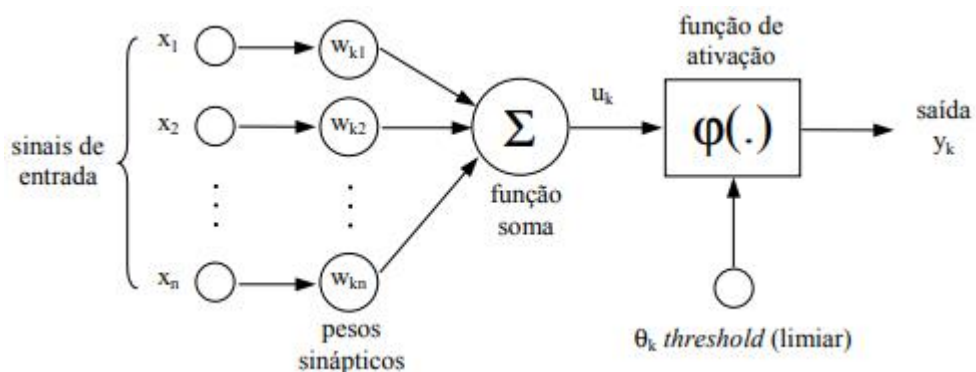


Figura 2. Modelo não linear de um neurônio segundo McCulloch e Pitts. Fonte: [29].

### 1.3 Clusterização

A Clusterização ou Agrupamento é um método utilizado para agrupar itens possuem características semelhantes, visa identificar e agrupar os registros similares. Um agrupamento (ou cluster) é formado por uma coleção de registros similares entre si, porém diferentes dos outros registros nos demais agrupamentos. A tarefa difere da classificação pois não necessita que os registros sejam previamente categorizados. Além disso, ela apenas identifica os grupos de dados similares, não tem o objetivo de classificar, estimar ou prever o valor de uma variável [14]. Na Figura 3 pode ser vista a identificação de clusters a partir de determinados registros.

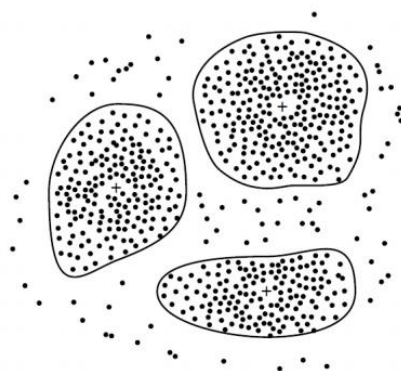


Figura 3. Registros agrupados em três clusters. Fonte: [14].

A tarefa de clusterização é mais utilizada durante a fase de preparação dos dados, além de geralmente utilizada junto a outras tarefas. As aplicações da clusterização são das mais variadas, pode-se citar pesquisas de mercado, reconhecimento de padrões, processamento de imagens, análise de dados, segmentação de mercado, pesquisas geográficas, classificação de documentos, detecção de comportamentos atípicos (fraudes), entre outras [15].

Dentre as técnicas de clusterização, uma das que mais se destaca é o algoritmo K-means. Trata-se de uma técnica de clusterização que busca minimizar a distância dos elementos a um conjunto de  $k$  centros de forma iterativa. Pode ser definido como uma técnica que utiliza o algoritmo de agrupamento de dados por K-médias. O seu objetivo é encontrar a melhor divisão de um número  $P$  de dados em  $K$  grupos  $C_i$ ,  $i = 1, \dots, K$ , de forma que a distância total entre os dados de um grupo e o seu centro, somada por todos os grupos, seja minimizada [5].

Esse algoritmo utiliza o conceito do centroide. Dado um conjunto de dados, o algoritmo seleciona de forma aleatória  $k$  registros, cada um representando um agrupamento. Para cada registro restante, é calculada a similaridade entre o registro analisado e o centro de cada agrupamento. O objeto é inserido no agrupamento com a menor distância, ou seja, maior similaridade. O centro do cluster é recalculado a cada novo elemento inserido [14]. A Figura 4 ilustra o funcionamento do algoritmo.

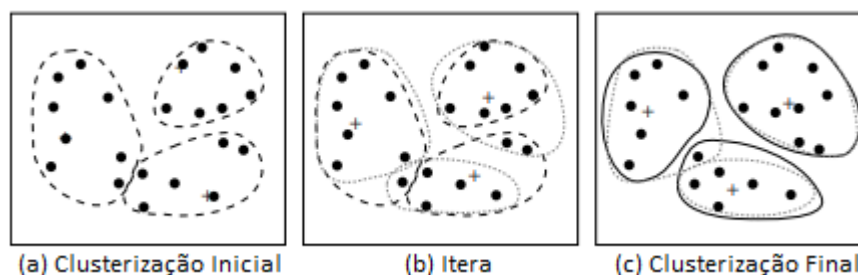


Figura 4. Clusterização de dados utilizando o algoritmo K-means. Fonte: [30].

Este algoritmo é extremamente veloz, geralmente convergindo em poucas iterações para uma configuração estável, na qual nenhum elemento está designado para um cluster cujo centro não lhe seja o mais próximo [16].

Um problema citado por [16] está relacionado ao fato do usuário definir a quantidade de clusters. No caso de uma má inicialização na quantidade de clusters poderá haver uma má separação dos dados. Um número muito

pequeno de conjuntos pode causar a união de dois ou mais clusters naturais, enquanto que um número muito grande pode fazer com que um cluster natural seja quebrado artificialmente em dois.

## 1.4 Rotas e Rastreamento Veicular

O termo rota é definido como percurso do ponto de partida até o ponto de destino [17]. Na área de gerenciamento de frotas o termo rota é muito utilizado para definir o percurso pelo qual o veículo percorreu ou deve ser percorrido.

Uma das formas mais comuns de gerenciamento de rotas se dá através do rastreamento veicular, com a instalação de um dispositivo de rastreamento no interior do veículo. Esse dispositivo possui um receptor GPS para a obtenção de informações a partir do recebimento de sinais enviados por satélites. O receptor tem por função realizar cálculos sobre as informações recebidas a fim de obter dados de localização como a latitude, longitude, altitude, data e hora. O envio das informações ocorre ao final do processo, sendo direcionadas a um centro de controle, que por meio de uma interface disponibiliza estas informações ao cliente [18].

A análise de rotas envolve um processo bastante repetitivo para os gestores de frotas que devem realizar uma análise bem criteriosa sobre as rotas percorridas pelos seus veículos, podendo em alguns casos não identificar uma ocorrência, ocasionando em perdas para a organização. Devido a este problema surge a necessidade de se desenvolver uma ferramenta de apoio a análise de rotas, facilitando assim o trabalho de análise dos gestores.

## 1.5 Trabalhos Relacionados

Nesta seção, com o objetivo de entender e agregar conhecimento foram analisados diversos trabalhos que envolvem inteligência artificial e a análise de rotas.

Laranjeiro et al. [19] utilizam em seu trabalho dados de rotas obtidos através de GPS para realizar uma ampla análise sobre a logística urbana na cidade de São Paulo. Os autores mostram como os dados de rotas podem ser utilizados para um melhor planejamento operacional das empresas e também de órgãos públicos, a fim de formular políticas de planejamento urbano mais adequadas.

Amita et al. [20] através de dados de GPS e o uso de RNA buscam prever o tempo de viagem do serviço de ônibus, com o objetivo de fornecer informações mais precisas sobre o tempo de viagem levando-se em consideração os diversos fatores que envolvem uma rota.

Vacca e Meloni [21] analisam em seu trabalho quais as características que mais influenciam o comportamento dos usuários na escolha de diferentes rotas para uma viagem que possui o mesmo par origem-destino.

S. Guo et al. [22] realizam um trabalho de segmentação de dados de rotas obtidas através de GPS baseado em lógica probabilística. Os autores fazem uma análise sobre as rotas geradas por veículos de transporte a fim de extrair informações de negócio avaliando o comportamento do veículo durante o trajeto.

Através da análise dos trabalhos relacionados constata-se que a principal contribuição deste trabalho se dá na identificação de rotas percorridas e na possibilidade de identificação de desvios de rotas, auxiliando assim, gestores de frotas no processo de análise de rotas.

# 2 Materiais e Métodos

## 2.1 Aquisição de Dados

Através de um sistema de rastreamento [23], foram coletados dados de rotas percorridas por veículos. Os dados coletados para o projeto consistem nos pontos (latitudes e longitudes) e a data e hora para ordenação dos pontos. Os pontos consistem nos locais onde o equipamento de rastreamento enviou sua localização após um tempo predefinido. O sistema de rastreamento utilizado tem por configuração padrão em seus equipamentos o envio da localização a cada sessenta segundos.

Os dados foram selecionados através do sistema e então tratados e convertidos para o formato de arquivo ARFF. O processo de tratamento das rotas é necessário para que as mesmas tenham um número padrão de pontos. Para isso, os autores desenvolveram uma ferramenta para a normalização das rotas, isto é, que faça a padronização da quantidade de pontos independente da distância, mantendo o percurso original da rota.

Para o desenvolvimento desta ferramenta foi utilizada a linguagem de programação Java. Seu funcionamento se dá na seguinte forma: - É fornecida uma rota de entrada (lista de pontos) e a quantidade de pontos que se deseja para a rota normalizada; - Realiza o cálculo da distância euclidiana entre cada um dos pontos a fim de obter a distância total da rota; - Utilizando a distância total da rota e o número de pontos desejado, calcula a distância média em que os pontos devem estar dispostos; - A partir do primeiro ponto da rota calcula qual será o próximo

ponto utilizando a distância média e a direção do segundo ponto da rota, e repete este processo até chegar ao último ponto; - Ao final da execução retorna uma lista de pontos com a quantidade de pontos solicitada.

### 2.1.1 Trajetos Utilizados

Foram definidos seis trajetos para servirem de base no trabalho de busca e análise dos resultados. Os trajetos foram extraídos de rotas reais percorridas por veículos rastreados pelo sistema Infratrack [23]. Abaixo pode-se verificar as características das rotas utilizadas:

- O primeiro trajeto se dá entre os municípios de Tapejara/RS e Soledade/RS, contando com uma distância de 56,4 quilômetros;
- O segundo trajeto se dá entre os municípios de Três de Maio/RS e Horizontina/RS, contando com uma distância de 23,8 quilômetros;
- O terceiro trajeto se dá entre os municípios de Horizontina/RS e Doutor Maurício Cardoso/RS, contando com uma distância de 18 quilômetros;
- O quarto trajeto se dá entre os municípios de Três de Maio/RS e São Martinho/RS, contando com uma distância de 35,3 quilômetros;
- O quinto trajeto se dá entre os municípios de Três de Maio/RS e Santa Rosa/RS, contando com uma distância de 33 quilômetros;
- O sexto trajeto se dá entre os municípios de Eldorado do Sul/RS e Cachoeirinha/RS, contando com uma distância de 23,8 quilômetros.

Na Figura 5 podem ser visualizadas as rotas utilizadas.

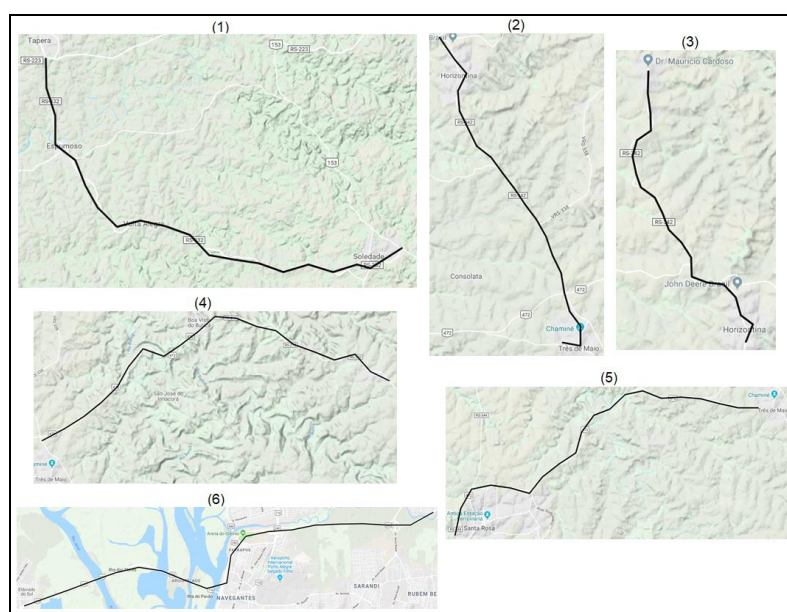


Figura 5. Rotas utilizadas no projeto final

## 2.2 WEKA

A ferramenta WEKA foi desenvolvida por pesquisadores da Universidade de Waikato, na Nova Zelândia, escrita em Java e sob a licença GPL. Fornece implementações de algoritmos de aprendizado que podem ser facilmente aplicados a conjuntos de dados. Inclui uma variedade de ferramentas para transformar conjuntos de dados, como os algoritmos para discretização e amostragem. Pode-se pré-processar um conjunto de dados, inseri-lo em um esquema de aprendizado e analisar o classificador resultante e seu desempenho, sem a necessidade de escrever nenhuma linha de código. O ambiente de trabalho inclui métodos para os principais problemas de mineração de dados: regressão, classificação, agrupamento, regras de associação e seleção de atributos. Todos os algoritmos recebem sua entrada na forma de uma única tabela relacional que pode ser lida de um arquivo ou gerada por uma consulta a um banco de dados [24].

## 2.2.1 Arquivo ARFF

Um arquivo ARFF (Formato de Arquivo Atributo-Relação) é um arquivo de texto escrito em formato ASCII. Ele descreve uma lista de instâncias que compartilham um conjunto de atributos [25]. Os arquivos ARFF foram desenvolvidos pelo projeto de Machine Learning do Departamento de Ciência da Computação da Universidade de Waikato, para uso no software Weka.

Os arquivos ARFF possuem duas seções. Sendo a primeira o Header, onde são definidos o nome da relação e a lista de atributos e seus tipos. A próxima seção é a Data que contém os dados, conforme descrito no Header [25]. Atualmente são suportados quatro tipos de dados: numérico, String (cadeia de caracteres), data e atributos nominais. Na Figura 6 pode-se visualizar um exemplo de arquivo ARFF.

```
% 1. Title: Iris Plants Database
%
% 2. Sources:
%   (a) Creator: R.A. Fisher
%   (b) Donor: Michael Marshall (MARSHALL%PLU@io.arc.nasa.gov)
%   (c) Date: July, 1988
@RELATION iris
@ATTRIBUTE sepallength NUMERIC
@ATTRIBUTE sepalwidth NUMERIC
@ATTRIBUTE petallength NUMERIC
@ATTRIBUTE petalwidth NUMERIC
@ATTRIBUTE class {Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa
4.4,2.9,1.4,0.2,Iris-setosa
4.9,3.1,1.5,0.1,Iris-setosa
```

Figura 6. Cabeçalho de arquivo ARFF. Fonte: [25].

## 2.3 Métodos de Clusterização e Aprendizado

### 2.3.1 Clusterização

Para o processo de clusterização foi utilizado o algoritmo K-means pela sua simplicidade e ótimos resultados na busca de padrões. Na Figura 7 pode ser visualizado o resultado de uma clusterização utilizando doze rotas de três trajetos diferentes. Pode-se visualizar que o algoritmo foi capaz de separar os dados em três clusters, cada um contendo quatro rotas. Os clusters são representados pelas cores azul, vermelho e verde.

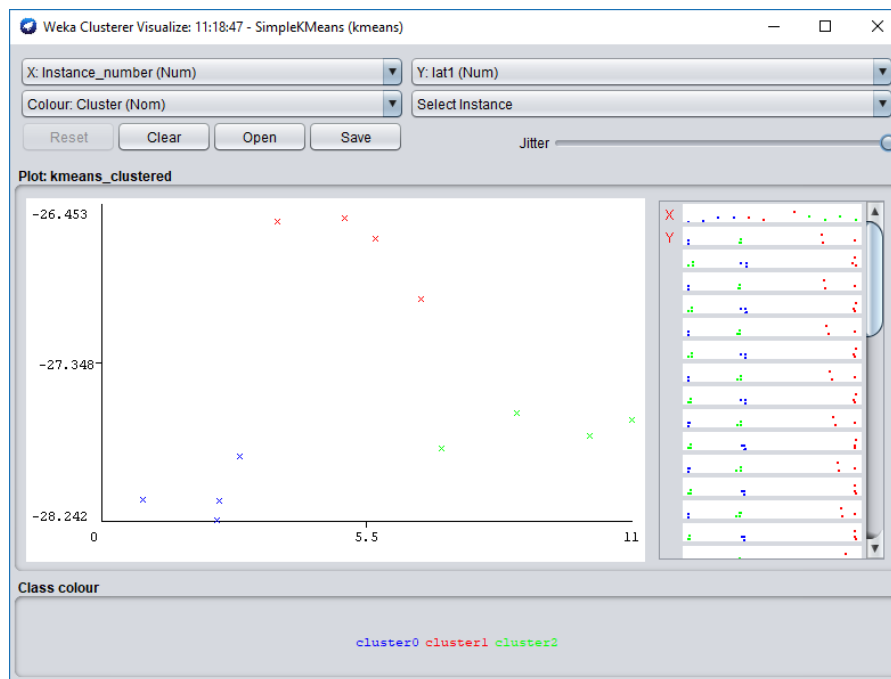


Figura 7. Visualização de clusters gerados pelo algoritmo K-means

### 2.3.2 Parâmetros Utilizados para a Clusterização

São disponibilizados no software Weka uma série de configurações a serem utilizadas para a aplicação dos algoritmos. No processo de escolha dos melhores parâmetros deve-se levar em consideração os dados utilizados e o problema em questão. Pode-se conferir na Tabela 1 os atributos, sua definição e os valores aplicados ao presente projeto. Os atributos relacionados ao Canopy tratam de um método de pré-clusterização não supervisionado muito utilizado em uma etapa de pré-processamento dos dados para o algoritmo K-means, reduzindo o número de cálculos de distância necessários [26].



Tabela 1: Parâmetros do algoritmo K-means no Weka

Atributo	Valor	Definição
canopyMaxNumCanopiesToHoldInMemory	100	Número máximo de Canopies a serem mantidos na memória durante o treinamento
canopyMinimumCanopyDensity	2.0	Densidade mínima baseada em T2, a partir da qual um Canopy será removido durante a remoção periódica
canopyPeriodicPruningRate	10000	Frequência com a qual se pode remover Canopies de baixa densidade durante o treinamento
canopyT1	-1.25	Distância T1 usada com a clusterização Canopy
canopyT2	-1.0	Distância T2 usada com a clusterização Canopy
distanceFunction	EuclideanDistance	Função de cálculo de distância
doNotCheckCapabilities	False	Os recursos do clusterizador não serão verificados antes que o clusterizador seja construído
dontReplaceMissingValues	False	Não substituir valores ausentes com média
fastDistanceCalc	False	Cálculo de distância mais rápido
initializationMethod	Farthest First	Método de inicialização
maxIterations	500	Número máximo de iterações
numClusters	3	Número de grupos a serem criados. Valor varia de acordo com as rotas apresentadas
numExecutionSlots	1	Número de threads para execução
preserveInstancesOrder	False	Preservar a ordem das instâncias
reduceNumberOfDistanceCalcsViaCanopies	False	Utilizar Canopies para reduzir o número de cálculos de distância realizados pelo K-means
seed	10	Número de sementes a serem utilizadas

### 2.3.3 Aprendizado

Para o processo de aprendizado foram utilizadas as Redes Neurais Artificiais. As RNAs possuem a capacidade de aprender sobre determinado problema a partir de informações do ambiente. Este aprendizado ocorre na fase de treinamento e é fundamental para o bom desempenho na próxima fase, conhecida como fase de validação ou de testes.

O treinamento do modelo e a validação são realizados através do método Treino-Teste. Nesse método 2/3 dos dados são destinados para o treinamento da rede neural e 1/3 dos dados são usados para validação do modelo [27]. Na fase de treinamento são inseridos dados conhecidos para o algoritmo, que fará a interpretação e criação de um modelo, o qual será utilizado na fase de testes. Após a criação do modelo, o mesmo é inserido no Weka junto de novos dados, desta vez com alguma informação incompleta, a qual será respondida pela RNA.

Na Figura 8 pode ser visualizado o resultado da fase de treinamento de um modelo.

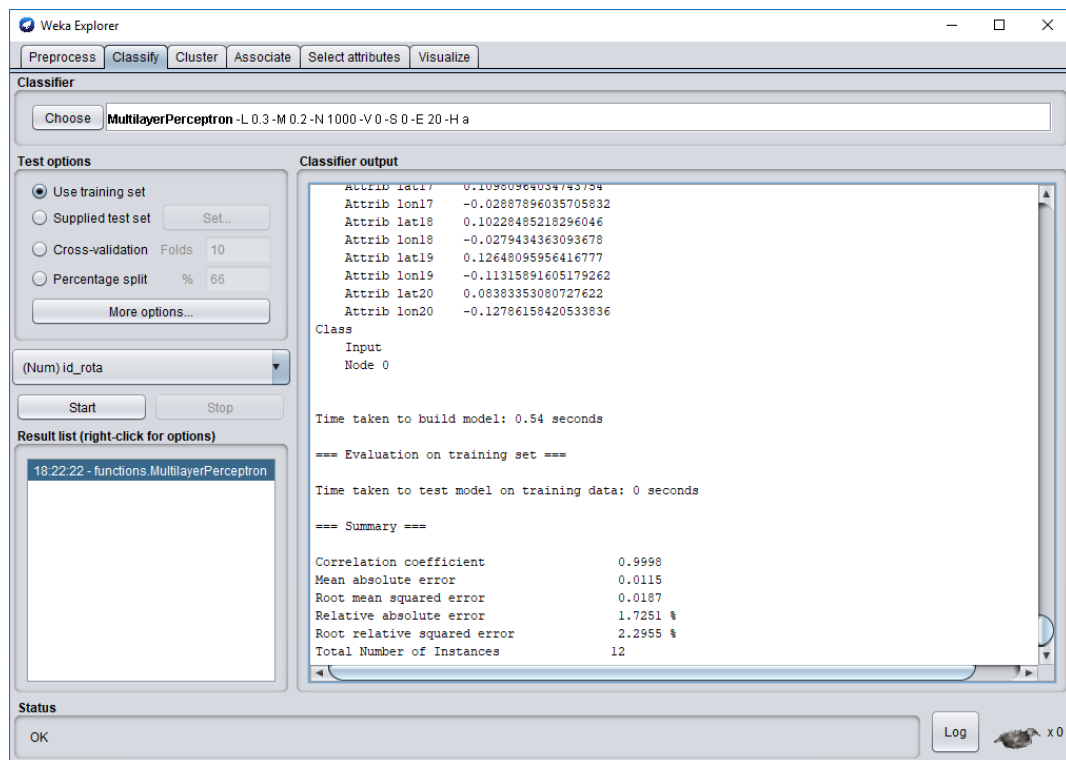


Figura 8. Resultado da fase de treinamento

Após o treinamento o Weka exibe uma série de informações referentes ao modelo criado, tais como: coeficiente de correlação, erro absoluto médio, erro quadrático médio, erro absoluto relativo, erro quadrático relativo e o número de instâncias utilizadas para o treinamento.

Após ter o modelo criado e salvo parte-se para a fase de testes para validação do modelo. Nesta fase é utilizado o modelo anteriormente criado e são selecionadas novas rotas, estas por sua vez não são identificadas para que a RNA possa encontrar o identificador correspondente a cada trajeto.

Conforme a Figura 9 o Weka exibe uma série de informações referentes ao resultado do teste. Na coluna predicted pode ser visualizado o resultado da predição da RNA. Fazendo-se o arredondamento dos números da coluna, obtém-se a rota a qual eles pertencem, pode-se observar que a linha 1 pertence a rota 1, já as linhas 2, 3 e 4 pertencem a rota 3 e a linha 5 não pertence a nenhuma rota. Os parâmetros utilizados durante a etapa de treinamento da RNA e seus respectivos valores são mostrados na Tabela 2.

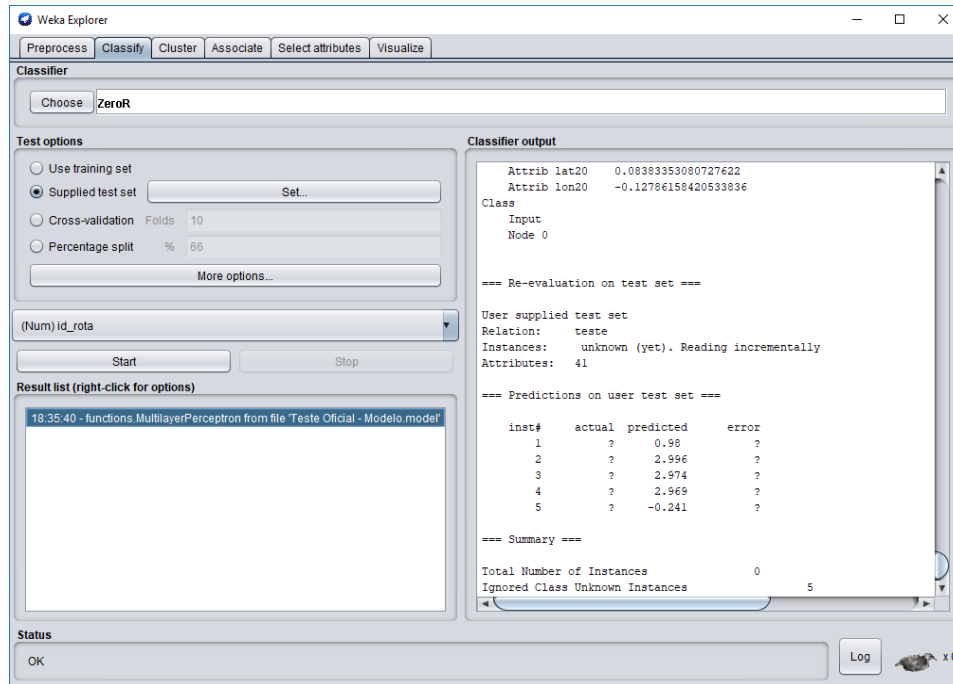


Figura 9. Resultado da fase de teste

### 2.3.4 Parâmetros Utilizados para o Treinamento do Modelo

Para o treinamento de um modelo de RNA é necessária a definição de uma série de parâmetros, os quais o Weka define um valor padrão que pode ser alterado conforme a necessidade. Estes parâmetros são utilizados somente na fase de treinamento da RNA, não sendo necessária a definição de nenhum parâmetro na fase de validação. A Tabela 2 mostra os atributos, sua definição e os valores utilizados para o treinamento do modelo.

Tabela 2: Parâmetros do algoritmo de RNA no Weka

Atributo	Valor	Definição
autoBuild	True	Adiciona e conecta camadas ocultas a rede
batchSize	100	Tamanho do batch padrão para previsões em lote
decay	False	Decair a taxa de aprendizado
doNotCheckCapabilities	False	Verifica as capacidades do associador antes que ele seja construído
hiddenLayers	10	Número de neurônios na camada oculta
learningRate	0.3	Taxa de aprendizado
momentum	0.2	Taxa de impulso
nominalToBinaryFilter	True	Dados devem ser convertidos de nominal para binário
normalizeAttributes	True	Não normalizar atributos
normalizeNumericClass	True	Não normalizar uma classe numérica
reset	True	Não é permitido redefinir a rede
seed	0	Valor usado para semear o gerador de números aleatórios
trainingTime	1000	Número de épocas do treinamento
validationSetSize	0	Porcentagem do tamanho do conjunto de validação a ser usado para finalizar o treinamento
validationThreshold	20	Limite de erros utilizado enquanto o teste de validação está sendo feito - Termina o teste quando o erro no conjunto de validação aumentou consecutivamente em um número predefinido de vezes

## 2.4 Simulações e Testes Utilizando RNAs

A fim de obter melhores resultados, foram realizadas simulações e testes utilizando diferentes parâmetros e configurações das RNAs e dados de rotas reais. Inicialmente foram selecionadas três rotas diferentes umas das outras. Após, devido à dificuldade na obtenção de novas amostras buscou-se três novas amostras para cada uma destas rotas a fim de serem utilizadas no treinamento da rede neural.

O arquivo de treinamento da RNA contou com doze amostras, cada uma delas identificada com o número correspondente a uma das rotas. Nos testes as rotas foram normalizadas para um número definido de pontos e inseridas no arquivo de treinamento.

Para o treinamento do modelo da RNA foram feitos testes utilizando diferentes configurações de RNAs, isto é, alterando-se o número de neurônios da camada de entrada e neurônios da camada oculta. Neste trabalho o número de neurônios da camada de entrada da RNA é dado pela equação:  $2 * nPontos + 1$  (dois multiplicado pelo número de pontos da rota mais um). Para rotas de 20 pontos são utilizados 41 neurônios: vinte latitudes, vinte longitudes e o código do trajeto correspondente. Já o número de neurônios da camada oculta é definido avaliando-se os resultados dos modelos criados.

Após, para a fase de validação são selecionadas novas rotas, sendo que todas as rotas utilizadas seguem o mesmo número de parâmetros das rotas utilizadas no treinamento. É alterado somente o valor do parâmetro de identificação da rota, onde é inserido o caractere “?” em seu lugar para que a RNA encontre o valor de identificação da rota.

## 3 Resultados e Discussão

Para obter e avaliar os resultados do desenvolvimento do projeto foi definido o seguinte método:

- Definir um conjunto de rotas a ser utilizado como base na exploração das ferramentas e métodos anteriormente definidos;
- Realizar simulações utilizando-se do conjunto de rotas a fim de verificar em quais hipóteses se obtém melhores resultados;
- Discutir e fundamentar a análise dos resultados obtidos.

### 3.1 Clusterização de Rotas

Utilizando o software Weka e o algoritmo K-means foi realizado o processo de clusterização das rotas. Foi utilizado o conjunto inicial de seis rotas e a coleta de novas amostras destas rotas, algumas destas amostras com desvios, outras não.

O processo de clusterização ou agrupamento apresentou bons resultados na identificação de rotas semelhantes. O algoritmo foi capaz de agrupar as rotas mais semelhantes em clusters iguais, inclusive nos casos de rotas com desvios, o algoritmo agrupa a rota desviada e a rota original correta.

No caso da descoberta e identificação dos desvios de rotas, o algoritmo faz de forma satisfatória a identificação dos desvios, desde que, o usuário insira a quantia correta de clusters que deseja. Desta forma, o uso do algoritmo K-means aplicado a descoberta de desvios de rotas não é muito indicado, já que, o usuário deverá saber previamente se no conjunto de rotas apresentado, alguma delas possui algum desvio, para aí o algoritmo separar esta rota em outro cluster.

Neste caso se forem apresentadas ao algoritmo uma rota correta e uma rota com desvio, o próprio usuário deverá informar que elas devem ser separadas em dois clusters.

Nas Figuras 10 e 11 pode-se verificar os testes feitos utilizando as rotas 4 e 5. Foram reunidas vinte e cinco amostras de cada uma das rotas e mais uma amostra com desvio de cada, totalizando cinquenta e duas amostras.

No caso do teste da Figura 10 define-se que o algoritmo deve encontrar dois clusters. O algoritmo efetua assim a separação das quatro amostras em dois conjuntos, sendo eles: rota 4 correta e amostra com desvio (em azul); rota 5 original e amostra com desvio (em vermelho).

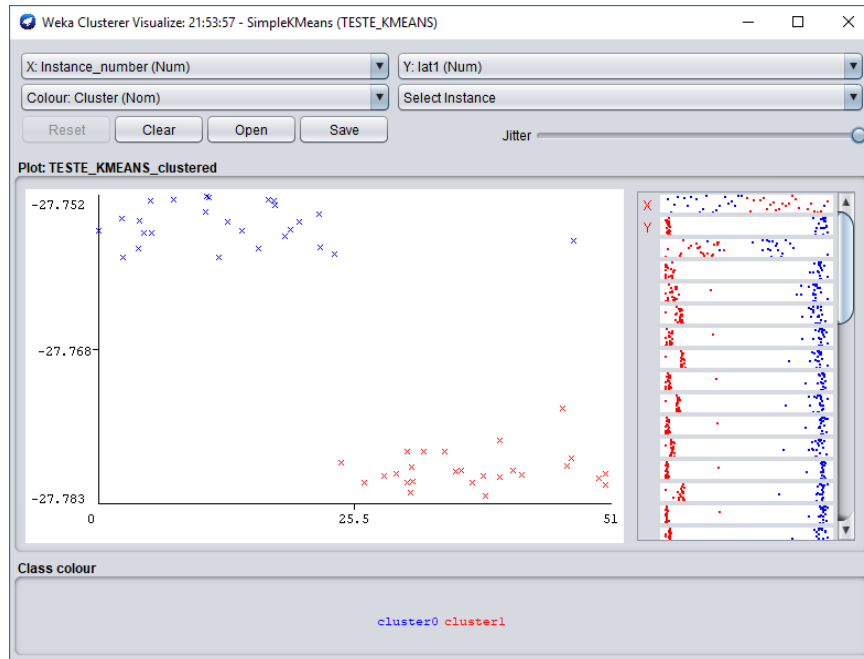


Figura 10. Separação em dois clusters

Na Figura 11 define-se que o algoritmo deve efetuar a separação das amostras em quatro clusters. Após o processamento ele separa as amostras em: rota 4 correta (em azul escuro); rota 4 com desvio (em azul claro); rota 5 correta (em vermelho); rota 5 com desvio (em verde). Reunindo assim, as rotas semelhantes em um cluster e separando a rota com desvio em outro cluster.

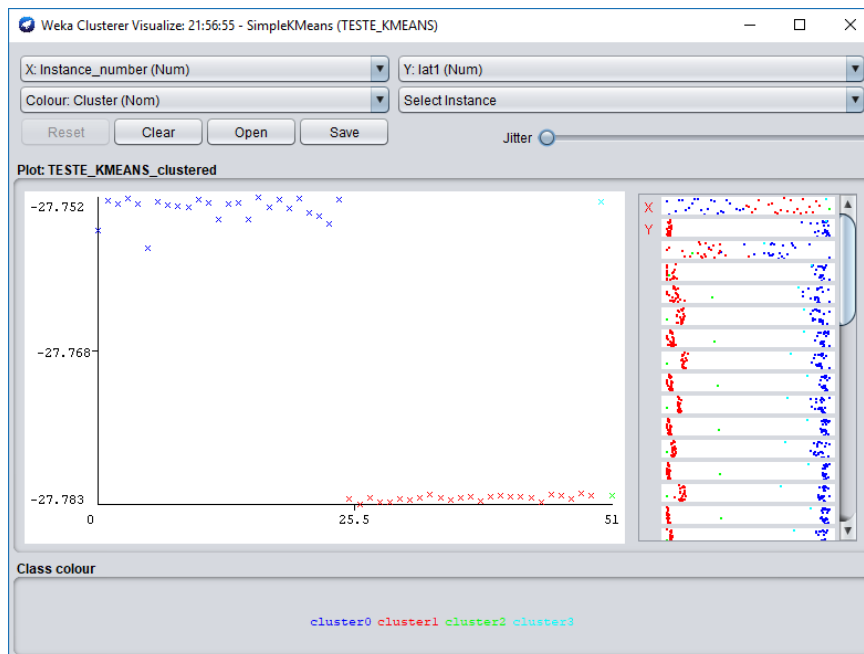


Figura 11. Separação em quatro clusters

### 3.2 Aprendizado de Rotas

Para o processo de aprendizado de rotas foi utilizado o algoritmo de Redes Neurais Artificiais.

### 3.2.1 Preparação dos Dados

Devido à dificuldade na extração das rotas foi definido um número reduzido de amostras. Para cada um dos trajetos foram coletadas vinte e cinco amostras, onde, o trajeto é o mesmo, porém com os pontos da rota posicionados em locais diferentes. A Figura 12 ilustra um exemplo, onde os pontos vermelhos pertencem a uma amostra de rota e os azuis a outra, sempre respeitando o trajeto utilizado como base.

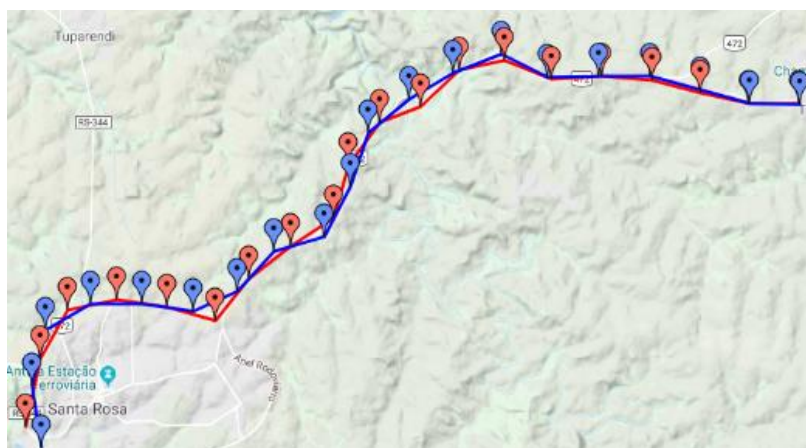


Figura 12. Exemplo de trajeto com duas amostras diferentes

Na próxima etapa os dados dos seis trajetos foram reunidos em um arquivo, totalizando cento e cinquenta amostras de dados. Para a identificação de cada amostra existem duas formas de se classificar os dados: numérico ou atributo nominal. Neste caso optou-se por identificar cada rota com valores numéricos, fazendo assim com que a RNA retorne um índice de semelhança com a rota definida, por exemplo: 1.02, 1.98 ou 3.87. No caso de se escolher um atributo nominal, a RNA retornaria com o número de identificação exato da rota apresentada, por exemplo: 1, 2 ou 4.

Foram realizados sucessivos testes para verificar qual o melhor método para ordenação dos dados. Na Tabela 3 pode-se visualizar os resultados obtidos com o treinamento da RNA em diferentes métodos de ordenação. Na coluna de Dados Agrupados os dados estavam agrupados por rotas, isto é, todas amostras da rota 1, todas amostras da rota 2, todas amostras da rota 3, e assim sucessivamente. Na coluna de Dados Sorteados os dados foram sorteados de forma aleatória. Já na coluna de Dados Intercalados os dados estavam intercalados, isto é, uma amostra da rota 1, uma amostra da rota 2, uma amostra da rota 3, outra amostra da rota 1, outra amostra da rota 2, e assim sucessivamente.

Tabela 3: Resultado do treinamento da RNA

Atributo	Dados Agrupados	Dados Sorteados	Dados Intercalados
Tempo Gasto para Construir o Modelo	0.21 segundos	0.23 segundos	0.27 segundos
Coefficiente de Correlação	0.992	0.9442	0.999
Erro Absoluto Médio	0.2047	0.398	0.0767
Erro Quadrático Médio	0.2507	0.5665	0.1053
Erro Absoluto Relativo	13.6481%	26.535%	5.1124%
Erro Quadrático Relativo	14.6777%	33.172%	6.1671%
Número de Instâncias	36	36	36

Através dos modelos criados, conclui-se que o melhor resultado foi obtido utilizando os dados das rotas de forma intercalada. Nos outros modelos, o percentual de erro apresentado foi superior, afetando diretamente a qualidade dos resultados das RNAs.

### 3.2.2 Treinamento

Para o treinamento da RNA foram utilizadas cento e oito amostras de dados. A fim de se encontrar a melhor configuração para a RNA, foram realizados vários testes utilizando diferentes números de neurônios na camada oculta. Para a escolha da melhor configuração foram observados valores como coeficiente de correlação e erro absoluto médio.

Após os testes concluiu-se que os melhores resultados foram encontrados utilizando-se 20 neurônios na camada oculta. Na Figura 13 pode-se visualizar o resultado exibido pelo Weka no treinamento da RNA.

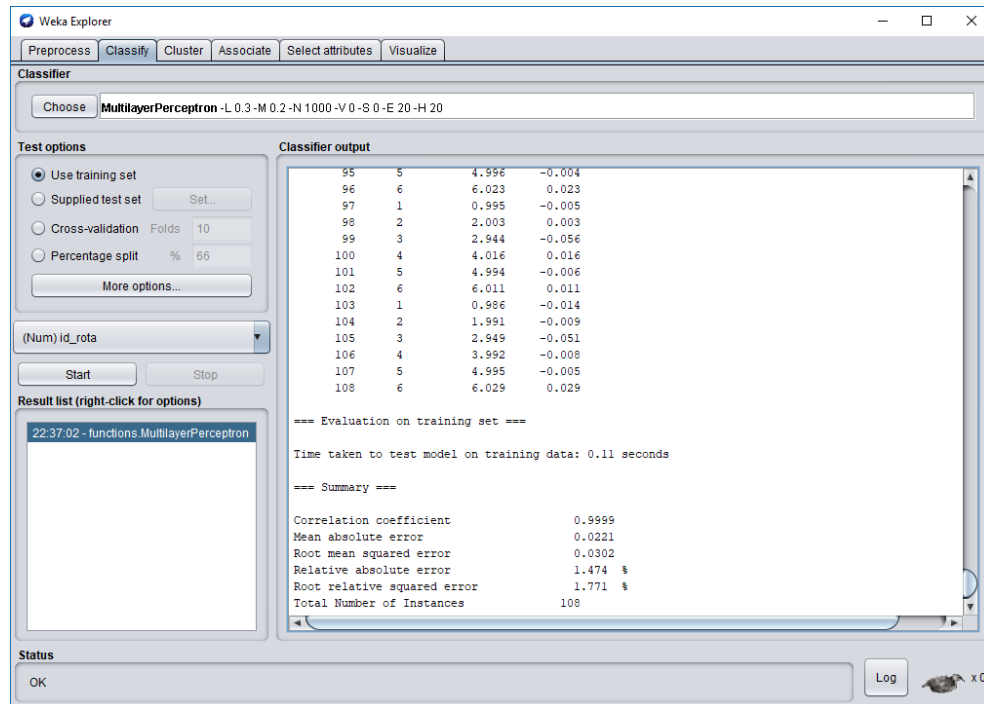


Figura 13. Resultado do treinamento da RNA

Os resultados exibidos pelo Weka podem ser visualizados na Tabela 4.

Tabela 4: Resultado do treinamento da RNA

Atributo	Valor
Tempo Gasto para Construir o Modelo	0.11 segundos
Coeficiente de Correlação	0.9999
Erro Absoluto Médio	0.0221
Erro Quadrático Médio	0.0302
Erro Absoluto Relativo	1.474%
Erro Quadrático Relativo	1.771%
Número de Instâncias	108

### 3.2.3 Validação

A partir do modelo criado iniciou-se os testes para validação. Para os testes foram utilizadas quarenta e duas amostras de dados. Neste teste não foram utilizadas rotas com desvio nem rotas inexistentes, a fim de avaliar a precisão dos resultados na identificação de rotas. Na Tabela 5 podem ser visualizados os resultados da predição das rotas feita pela RNA. Na Tabela (5) constam a ordem das instâncias apresentadas ao Weka, a identificação da rota e o resultado da predição.

Tabela 5: Resultado da predição de rotas

Inst#	Rota Apresentada	Predição
1	1	0.979
2	1	0.968
3	1	0.986
4	1	0.986
5	1	0.974
6	1	0.998
7	1	0.98
8	2	1.905
9	2	1.954
10	2	1.936
11	2	1.943
12	2	2.02
13	2	1.981
14	2	2.004
15	3	2.959
16	3	2.992
17	3	2.958
18	3	2.931
19	3	2.941
20	3	2.944
21	3	2.945
22	4	4.002
23	4	3.995
24	4	4.009
25	4	3.988
26	4	3.992
27	4	4.01
28	4	4.007
29	5	5.01
30	5	4.983
31	5	4.98
32	5	4.966
33	5	4.963
34	5	5.011
35	5	5
36	6	6.014
37	6	6.008
38	6	6.043
39	6	6.036
40	6	6.021
41	6	5.973
42	6	6.004

Desta forma, no quesito identificação de rotas, utilizando-se da técnica de arredondamento dos resultados, a RNA apresentou uma taxa de acerto de 100%. Possuindo uma variação máxima de 0.095 para mais ou para menos em relação ao resultado correto.

### 3.3 Identificação de Desvios de Rotas

Na identificação de desvios de rota optou-se por utilizar somente as RNAs devido a forma com que estas fazem a identificação das rotas. Foi criado um novo modelo, o qual foi criado utilizando-se cento e cinquenta amostras de



dados e os mesmos parâmetros definidos no modelo anterior. Para a validação do modelo foram escolhidas três novas amostras de cada um dos trajetos com diferentes níveis de desvios de rota.

Na Tabela 6 podem ser visualizados os resultados da predição dos desvios de rota. Na Tabela (6) constam a ordem das instâncias apresentadas ao Weka, a identificação da rota, o resultado da predição e o cálculo da porcentagem de desvio, que é baseado na seguinte equação:  $(\text{Diferença entre Predição e Rota Apresentada} / \text{Valor de Variação Máxima}) * 100$ . O Valor de Variação Máxima pode ser ajustado, para o teste foi definido o valor de 0.5.

Tabela 6: Resultado da predição de desvios de rota

Inst#	Rota Apresentada	Predição	% de Desvio
1	1	0,517	97%
2	1	0,737	53%
3	1	1,616	123%
4	2	2,408	82%
5	2	2,291	58%
6	2	1,904	19%
7	3	3,256	51%
8	3	3,287	57%
9	3	3,623	125%
10	4	3,464	107%
11	4	3,820	36%
12	4	4,171	34%
13	5	3,808	238%
14	5	3,670	266%
15	5	4,401	120%
16	6	5,593	81%
17	6	5,469	106%
18	6	6,239	48%

No quesito de identificação de desvios de rota, ignorando-se os valores onde o desvio foi superior a 100%, a RNA foi capaz de obter um índice de desvio em 11 dos 18 desvios de rota. Nos casos onde o desvio foi superior a 100% espera-se que com um treinamento melhorado da RNA, isto é, com um conjunto maior de dados, a RNA possa obter um índice de desvio.

### 3.4 Comparação Entre Clusterização e Aprendizado

Comparando-se o algoritmo K-means e as RNAs observa-se que o algoritmo K-means é uma excelente ferramenta aplicada na identificação e agrupamento das rotas mais semelhantes. No caso da identificação de desvios de rotas o problema do algoritmo consiste no usuário informar o número de grupos que ele deseja. O algoritmo só fará corretamente a separação de um desvio se o usuário souber e informar o número correto de grupos a ser gerado. Por exemplo, se for apresentado um conjunto de cinco dados da mesma rota e um destes dados possuir um desvio de rota, o próprio usuário deverá informar ao algoritmo que ele deve separar os dados em dois grupos, um com as quatro rotas semelhantes e outro com a rota de desvio. Desta forma o algoritmo não se aplica a descoberta de desvios de rotas, somente na identificação mediante descoberta.

Utilizando as RNAs no problema proposto de identificação de rotas e de desvios de rotas, as RNAs apresentaram excelentes resultados na identificação das rotas e resultados satisfatórios na identificação dos desvios de rotas. Com este método é possível descobrir a qual rota determinado dado pertence, além de obter um índice de semelhança com a rota que pode ser aplicado na detecção de um desvio de rota.

As RNAs se comparadas ao algoritmo K-means tem como ponto negativo o trabalho necessário na coleta de mais amostras, no treinamento e validação do modelo, enquanto que no K-means é necessário um conjunto menor de dados e os dados só precisam ser apresentados ao algoritmo, não sendo necessária nenhuma fase adicional.

## 4 Conclusão

Este trabalho se propôs a explorar e demonstrar a viabilidade da utilização de técnicas de Inteligência Artificial, como o algoritmo K-means e as Redes Neurais Artificiais em problemas de rotas. Após a análise teórica,

desenvolvimento, simulações e testes, os resultados obtidos foram considerados satisfatórios. Sendo assim, um trabalho como este pode ser aproveitado para o auxílio de especialistas da área e ser relevante em diversas áreas relacionadas a roteirização de veículos.

Comparando-se os algoritmos K-means e as RNAs, ambos apresentam pontos positivos, sendo as RNAs um método mais adequado ao problema proposto. Através dos testes foi possível concluir que elas são um excelente método de identificação de rotas, já o algoritmo K-means apresentou excelentes resultados de identificação de rotas mais semelhantes.

Referente aos desvios de rotas, as RNAs distinguem muito bem desvios maiores de rotas, porém apresentam dificuldades na identificação de pequenos desvios. Acredita-se que este problema possa ser contornado utilizando-se uma concentração maior de pontos nas rotas, tornando-as mais precisas para a RNA.

Definir a quantidade de pontos das rotas mais adequado é o desafio das RNAs, levando-se em consideração que as rotas podem ser de curta, média ou longa distância e de que para o treinamento e validação as RNAs precisam de um número fixo de parâmetros.

Conclui-se que para a identificação de rotas as RNAs são um excelente método, capaz de auxiliar gestores de frotas no controle sobre as rotas percorridas por seus veículos, inclusive nos casos de desvios de rotas, apesar de possuir uma taxa de acerto menor neste caso. Utilizando-se de uma ferramenta como esta é possível otimizar tempo além de evitar o serviço repetitivo de verificação do trajeto percorrido pelos veículos.

Como trabalhos futuros sugere-se realizar testes mais aprofundados utilizando as RNAs. Como por exemplo avaliar se com o aumento da quantidade de pontos nas rotas é possível melhorar o resultado na identificação de desvios, além de testes utilizando um modelo de RNA específico para cada rota a fim de obter melhores resultados na identificação de desvios de rotas.

## Referências

- [1] SELLITTO, M. Inteligência artificial: uma aplicação em uma indústria de processo contínuo. *Gestão & Produção*, v. 9, n. 3, p. 363-376, 2002. DOI: <https://dx.doi.org/10.1590/S0104-530X2002000300010>.
- [2] ARAUJO, Jefferson Borges, MACHADO, Viviane da Silva. Logística e Atualidade - Empregando TI em Novas Oportunidades. *Revista Uniabeu*, v. 4, n. 7, p. 122-133, 2011.
- [3] SANTOS, Alysson Vinícius Neves dos; FELIX, Leonardo Bonato; VIEIRA, José Geraldo Vidal. Estudo da logística de distribuição física de um laticínio utilizando lógica fuzzy. *Produção*. v. 22, n. 3, p. 576-583, 2012. DOI: <https://dx.doi.org/10.1590/S0103-65132012005000036>.
- [4] CORRÊA, Fernanda. Aplicação de redes neurais artificiais no setor de transportes no Brasil. 2008. 72 f. Dissertação (Mestrado em Engenharia Urbana) – Universidade Federal de São Carlos, São Carlos, 2008.
- [5] PIMENTEL, Edson P.; DE FRANÇA, Vilma F.; OMAR, Nizam. A identificação de grupos de aprendizes no ensino presencial utilizando técnicas de clusterização. In: *Simpósio Brasileiro de Informática na Educação - SBIE*. 2003. p. 495-504. DOI: <https://dx.doi.org/10.5753/cbie.sbie.2003.495-504>.
- [6] SCHILDT, H. Inteligência Artificial Utilizando Linguagem C. São Paulo: McGrawHill, 1989. 349 p.
- [7] CERON, Cristiano; SCHÜTZ, Sérgio Murilo. Sistema Especialista de Apoio para Controle de Pragas e Doenças no Cultivo da Maçã. II Workshop de Tecnologia da Informação aplicada ao Meio Ambiente – CBComp. p. 737-740, 2004.
- [8] JESS, Gil Marcos. Inteligência artificial e tecnologias da inteligência – Um repensar segundo os processos de elaboração matemática. 2004. 139 f. Dissertação (Mestrado em Educação Matemática) – Universidade Federal do Paraná, Curitiba, 2004.
- [9] DOUGHERTY, Mark. A review of neural networks applied to transport. *Transportation Research Part C: Emerging Technologies*, v. 3, n. 4, p. 247-260, 1995. DOI: [https://doi.org/10.1016/0968-090X\(95\)00009-8](https://doi.org/10.1016/0968-090X(95)00009-8).
- [10] MENDES FILHO, Elson Felix. Projeto evolucionário de redes neurais artificiais para avaliação de crédito financeiro. 1997. 96 f. Dissertação (Mestrado em Ciências – Área: Ciências de Computação e Matemática Computacional) – Universidade de São Paulo, São Carlos, 1997.
- [11] BARRETO, Jorge M. Introdução às redes neurais artificiais. V Escola Regional de Informática. Sociedade Brasileira de Computação, Regional Sul, Santa Maria, Florianópolis, Maringá, p. 5-10, 2002.

- [12] COSTA, Paulo Teixeira da. Uma análise do consumo de energia em transporte nas cidades portuguesas utilizando Redes Neurais Artificiais. 2003. 146 f. Dissertação (Mestrado em Engenharia Municipal) – Universidade do Minho, Braga, 2003.
- [13] RAIÁ JUNIOR, Archimedes Azevedo. Acessibilidade e mobilidade na estimativa de um índice de potencial de viagens utilizando redes neurais artificiais e sistemas de informações geográficas. 2000. 196 f. Tese (Doutorado em Engenharia Civil - Transportes) – Universidade de São Paulo, São Carlos, 2000.
- [14] CAMILO, Cássio Oliveira; SILVA, João Carlos da. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. Universidade Federal de Goiás (UFG), p. 1-29, 2009.
- [15] OLIVEIRA, R. R.; CARVALHO, C. L. Algoritmos de agrupamento e suas aplicações. Universidade Federal de Goiás (INF-UFG), 2008.
- [16] LINDEN, Ricardo. Técnicas de agrupamento. Revista de Sistemas de Informação da FSMA, v. 4, p. 18-36, 2009.
- [17] BORGES JÚNIOR, Altamiro C.; GALVÃO, Alex. Dicionário de Logística. [201-?]. Disponível em: <<http://www.tecnologistica.com.br/portal/dicionario/logistica>>. Acesso em: 06 Jun. 2018.
- [18] PRETI, Fábio Marin. Automação de rastreamento veicular on-line para análise de risco. 2011. 118 f. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Itajubá, Itajubá, 2011.
- [19] LARANJEIRO, Patricia F. et al. Using GPS data to explore speed patterns and temporal fluctuations in urban logistics: The case of São Paulo. *Journal of Transport Geography*, v. 76, p. 114-129, 2019. DOI: <https://doi.org/10.1016/J.JTRANGE.2019.03.003>.
- [20] AMITA, Johar; JAIN, S.S.; GARG, P.K. Prediction of Bus Travel Time Using ANN: A Case Study in Delhi. *Transportation Research Procedia*, v. 17, p. 263-272, 2016. DOI: <https://doi.org/10.1016/j.trpro.2016.11.091>.
- [21] VACCA, Alessandro; MELONI, Italo. Understanding Route Switch Behavior: An Analysis Using GPS Based Data. *Transportation Research Procedia*, v. 5, p. 56-65, 2015. DOI: <https://doi.org/10.1016/j.trpro.2015.01.018>.
- [22] GUO, Sini et al. GPS trajectory data segmentation based on probabilistic logic. *International Journal of Approximate Reasoning*, v. 103, p. 227-247, 2018. DOI: <https://doi.org/10.1016/j.ijar.2018.09.008>.
- [23] INFRATRACK. Site Infratrack, 2019. Disponível em: <<https://www.infratrack.com.br>>. Acesso em: 10 Mai. 2019.
- [24] WITTEN, I. H. et al. *Data Mining: Practical Machine Learning Tools and Techniques*. 4. ed. Morgan Kaufmann, 2016. 654 p. DOI: <https://doi.org/10.1016/C2015-0-02071-8>.
- [25] WAIKATO. Attribute-Relation File Format (ARFF). Disponível em: <<https://www.cs.waikato.ac.nz/ml/weka/arff.html>>. Acesso em: 20 Out. 2018.
- [26] MCCALLUM, Andrew; NIGAM, Kamal; UNGAR, Lyle H. Efficient clustering of high-dimensional data sets with application to reference matching. In: *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2000. p. 169-178. DOI: <https://doi.org/10.1145/347090.347123>.
- [27] OLSON, David L.; DELEN, Dursun. *Advanced Data Mining Techniques*. Editora Springer, 2008. DOI: <https://doi.org/10.1007/978-3-540-76917-0>.
- [28] SCHUTZER, Daniel. *Artificial intelligence: an applications-oriented approach*. New York: Van Nostrand Reinhold Company, 1987. DOI: <https://doi.org/10.1177/089443938900700230>.
- [29] HAYKIN, Simon. *Neural Networks: A comprehensive foundation*. 2. ed. Hamilton: Pearson Education, 1994. 823 p.
- [30] HAN, Jiawei; PEI, Jian; KAMBER, Micheline. *Data mining: concepts and techniques*. Elsevier, 2011.