

# TÉCNICAS DE MINERAÇÃO DE DADOS PARA PREDIÇÃO DE CONSUMO ENERGÉTICO DE UM VEÍCULO ELÉTRICO

GIULIANNA N. T. MANDOLESI<sup>1</sup>, STANLEY R. DE M. OLIVEIRA<sup>2</sup>

## RESUMO:

Este trabalho descreve o desenvolvimento de técnicas de mineração de dados para que fosse possível a predição do consumo energético de um veículo elétrico em um percurso, bem como a sua autonomia. Foram considerados 5 percursos distintos de um carro do tipo fórmula, com mais de 90 dados distintos aquisitados durante a rota. Uma vez que a quantidade de dados por percurso é muito maior que a quantidade de percurso, foi necessária a aplicação de um método de seleção de atributos, o *Principal Component Analysis* (PCA). Também utilizando os algoritmos K-Nearest Neighbors e K-means foram removidos ruídos dos dados.

Além disso, foram consideradas situações distintas de temperatura ambiente, motorista e torque máximo do carro. Os estudos mostraram que a temperatura não exerceu grande influência sobre a energia gasta, diferente do tempo de volta e da velocidade máxima atingidas, que tiveram maior influência. Prevê-se que os modelos desenvolvidos possam ser utilizados para estimar o consumo total de energia nos estágios iniciais do projeto quando é dimensionado o pack de baterias de um veículo.

**PALAVRAS-CHAVE:** *predição de energia; seleção de atributos; data mining*

## DATA MINING TECHNIQUES FOR PREDICTION OF ENERGY CONSUMPTION OF AN ELECTRIC VEHICLE

## ABSTRACT:

The present study describes the development of data mining techniques to predict the energy consumption of an electric vehicle in a route, as well as its autonomy. Five different routes of a formula-type car were considered, with more than 90 different data acquired during the route. Since the amount of data per path is much greater than the amount of path, it was necessary to apply an attribute selection method, the *Principal Component Analysis* (PCA). Also, using K-Nearest Neighbors and K-means algorithms, noises of the data were removed.

In addition, different situations of ambient temperature, driver and maximum torque of the car were considered. The studies showed that the temperature had no influence on the energy expended, other than the time of return and the maximum speed reached, that had greater influence. It is envisioned that the models developed can be used to estimate the total energy consumption in the initial stages of the project when the battery pack of a vehicle is sized.

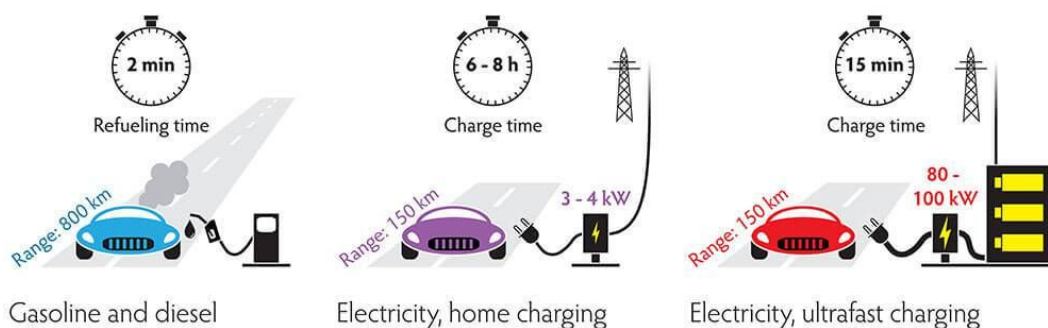
**KEYWORDS:** *energy prediction; selection of attributes; data mining*

<sup>1</sup> Graduanda em Eng. Elétrica, UNICAMP/Campinas, SP - [giulianna.mandolesi@gmail.com](mailto:giulianna.mandolesi@gmail.com)

<sup>2</sup> Doutor em Ciencia da Computacao, Pesquisador da Embrapa Informatica Agropecuaria, Professor do Programa de Pos-Graduação da Feagri, UNICAMP - [stanley.oliveira@embrapa.br](mailto:stanley.oliveira@embrapa.br)

## INTRODUÇÃO

A sustentabilidade dos setores de energia e de transporte provida pela mobilidade elétrica tem sido determinante para a sua expansão. Contudo, ela apresenta grandes limitações quando se trata da capacidade de armazenamento de baterias e de sua carga. Quando comparada com veículos a combustão, um veículo elétrico demora cerca de 7x mais para ser reabastecido, em uma carga super rápida, para atingir uma autonomia significativamente menor, conforme apresentado na Figura 1.



**Figura 1.** Ilustração de Reabastecimento para **a)** Carros Combustão. **b)** Carros elétricos carregados em casa. **c)** Carros elétricos em postos de carga super rápida

Assim, uma predição de autonomia precisa se torna necessária para que o piloto saiba se chegará em seu destino sem que a bateria se torne um limitante, seja numa viagem, seja para uma reunião de negócios. O impacto também pode ser visto na indústria: montadoras poderão, baseadas nesta predição, otimizar o projeto do pack de baterias em relação ao desejado visto o perfil do cliente, podendo, inclusive, reduzir custos.

Neste mesmo contexto de desenvolvimento da tecnologia de mobilidade elétrica, a equipe Unicamp E-Racing desenvolve um carro do tipo Formula elétrico. Este se diferencia dos carros comerciais nos mais diversos aspectos: é um monoposto, possui cerca de 20% do peso dos veículos comuns e o seu ambiente é muito mais hostil, com desacelerações e acelerações abruptas, pouco vistas no dia-a-dia. Contudo, por ser treinado em uma pista fechada, os seus dados apresentam uma periodicidade bastante nítida, o que será essencial na implementação dos algoritmos.

Sendo assim, os dados serão primeiramente divididos por volta, a partir de técnicas de Mineração de Dados para a determinação de cada uma das voltas. Feito, isso, os dados terão que ser selecionados para a predição do consumo energético, uma vez que são mais de 90 atributos descrevendo cada um dos trajetos. Contudo, a influência deles na autonomia de um carro é um setor ainda pouco explorado.

Desta forma, este desafio alinha-se à proposta da Mineração de Dados de estimar modelos a partir de relações e padrões ainda desconhecidos. No desenvolver deste projeto, serão então apresentadas algumas de suas técnicas para a predição do consumo energético de um veículo elétrico.

## MATERIAL E MÉTODOS

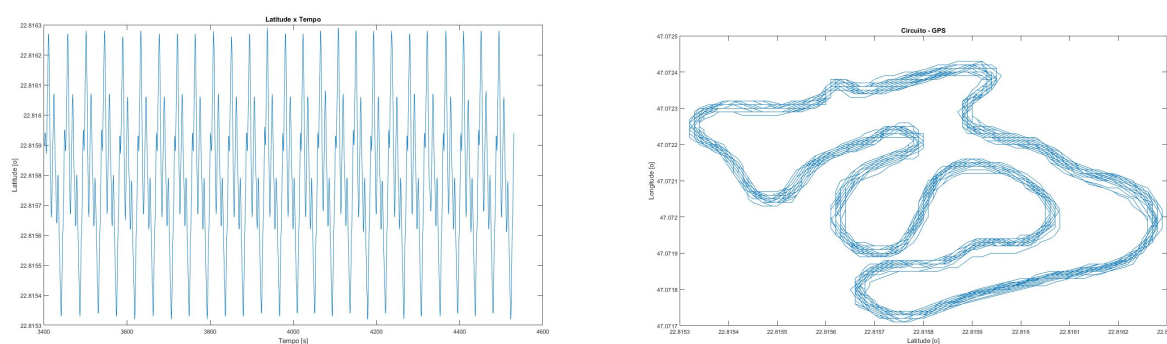
### Base de Dados de um Veículo Elétrico

Na primeira etapa, foi utilizada uma base dados fornecida pela equipe Unicamp E-Racing, a qual desenvolve um monoposto do tipo elétrico para competições estudantis de Formula SAE. Esta base é composta por dados de 5 testes, cada um realizado em pistas e condições atmosféricas distintas e com o registro de 92 atributos numéricos distintos, relacionados principalmente ao motor, bateria e performance dinâmica do carro.

### Limpeza dos Dados

Uma vez que nem todos os elementos são necessários para o problema em questão, fez-se necessária a seleção de atributos pertinentes, além da remoção de ruídos, a fim de aumentar a qualidade dos dados. Esta se trata da segunda etapa, de preparação de dados, e se mostra de suma importância, uma vez que melhora fortemente a precisão do modelo, além de reduzir a sua complexidade. Para isso, foram implementados algoritmos em Matlab com o auxílio das *Toolboxes Statistics and Machine Learning* e *Signal Processing*.

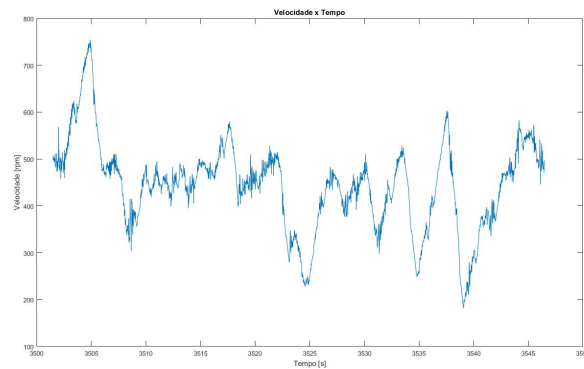
Primeiramente baseado em conhecimentos prévios, foram selecionados os dados pertinentes para cada um dos casos. Para reconhecer em qual volta o piloto se encontra, os dados do GPS, apresentados na Figura 2a e 2b, se apresentam suficientes. É visível não só uma periodicidade do sinal, mas também um certo ruído nos dados, que pode ser atribuído a um acúmulo de erros no sensor, e também a um deslocamento do veículo na pista (ora o piloto faz uma curva mais fechada, ora mais aberta).



**Figura 2.** a) Gráfico de Latitude em Função do Tempo. b) Mapeamento da Pista a partir de dados do GPS

O algoritmo de K-means foi escolhido então para suavizar o ruído dos dados. Apesar de este ser utilizado geralmente como classificador, parte da sua lógica foi utilizada para então esta implementação: foram determinadas regiões a partir da densidade de dados e encontrado o centro de massa para cada uma delas. Este se tornou então o ponto na pista em questão.

Então, para prever a energia a ser gasta, foram considerados os atributos, de cada uma das voltas: Energia, Tempo de Volta, Velocidade Máxima, Torque Máximo, Torque Regenerativo Máximo, Demanda de Pedal Média e as Temperaturas Médias do Arrefecimento do Inversor, do Rotor e do Estator do Motor e das Baterias. Mais uma vez, tornou-se necessário pré-processar os dados, uma vez que os dados apresentaram grandes níveis de ruído, como a velocidade do veículo, apresentada na Figura 3.



**Figura3.** Velocidade do veículo em uma Volta

A técnica de K-Nearest Neighbors foi então escolhida para isso. Esta técnica consiste em observar o comportamento dos  $k$  vizinhos de uma variável para determinar o valor da mesma. Apesar de esta ser usualmente utilizada para determinar valores faltantes ela foi, neste caso, utilizada para determinar qual seria o ponto de cada intervalo de tempo, suavizando o ruído dos dados.

Então, utilizando a Análise de Componentes Principais (PCA), que consiste na ortogonalização de vetores a fim de encontrar a correlação de variáveis e, assim, selecionar os atributos mais pertinentes para prever a energia gasta, simplificando o modelo final a ser desenvolvido.

### **Determinação de Voltas Dadas**

Tendo em vista que, a cada volta dada em uma pista, o comportamento espacial do carro é aproximadamente periódico, foi implementado um algoritmo que encontrasse a similaridade entre dois sinais: a partir da primeira volta, o algoritmo deveria ser, por si só, determinar qual parte do dado fazia parte de qual volta. Para isto, foi utilizado inicialmente a medida de correlação cruzada, que encontra em quais instantes de tempo o padrão da primeira volta se repete ao longo do sinal.

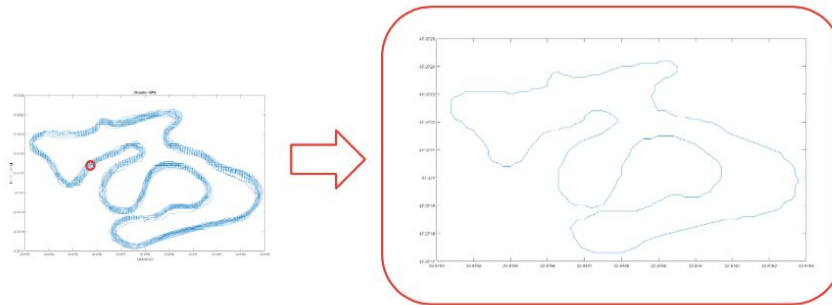
### **Predição de Energia a ser Gasta**

Com os dados preparados e devidamente selecionados, foi então implementada a técnica de Regressão Múltipla. Esta, além de possuir um custo computacional baixo, permite a combinação de diversos fatores para a determinação do valores final. Dado o valor de energia para a volta, é possível prever a autonomia, uma vez que a capacidade energética do banco de baterias é conhecido. Analogamente, é possível também determinar qual deve ser a capacidade energética do banco de baterias uma vez que o gasto energético é conhecido, sendo esta uma aplicação de bastante interesse na indústria.

## RESULTADOS E DISCUSSÃO

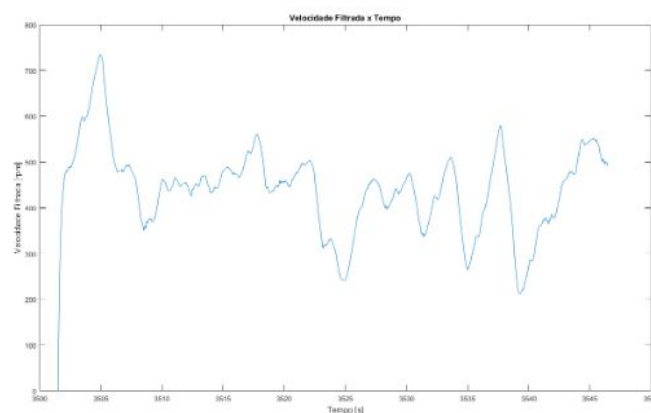
### Limpeza dos Dados

Utilizando o Matlab, foi implementado um algoritmo de K-means que definia clusters nas regiões de maior densidade do mapeamento da pista. Então, considerando que os dados iniciais possuem menos erro oriundo do acúmulo de erro do sensor inercial, foi dado um peso para cada um dos pontos daquele cluster, baseado na volta em que foi obtido, a fim de penalizar aquelas com maior acúmulo de erro. Também foi considerado que, na primeira volta, o piloto está reconhecendo a pista e, então, também deve ser penalizada. Considerando  $n$  voltas, à primeira, foi atribuído um peso de 1.5; da segunda à  $n/2$ -ésima volta, 2; e para o restante, o peso de 1. O resultado obtido segue apresentado na Figura 4.



**Figura 4.** Limpeza do Mapeamento da Pista

Então, com o algoritmo de KNN também implementado em Matlab foram filtrados os dados de Corrente, Torque e Velocidade do carro, todas ruidosas. A técnica implementada foi eliminar temporariamente o dado de um determinado instante de tempo e determinar o seu valor a partir de 6 dos seus vizinhos mais próximos para a Corrente e 10 para o Torque e a Velocidade do Carro. Este valor foi calibrado observando o gráfico final obtido em relação ao original. A grande vantagem desta implementação é que, diferente de filtros passa-baixa, por exemplo, este não apresentou um atraso no sinal, ao mesmo tempo que suavizou o ruído. O resultado final para a Velocidade segue apresentado na Figura 5 como exemplo.



**Figura 5.** Curva de Velocidade no Tempo com Filtro pelo KNN

Com o ruído eliminado, puderam ser extraídas informações relevantes sobre cada um dos atributos, apresentadas nas Tabela 1a e 1b.

**Tabela 1a.** Dados processados por Volta para Predição de Energia

<b>Volta</b>	<b>Energia (kW)</b>	<b>Tempo de Volta (s)</b>	<b>Velocidade Máxima (rpm)</b>	<b>Torque Máximo (Nm)</b>	<b>Torque Regenerativo Máximo (Nm)</b>	<b>Demanda Média de Pedal (%)</b>
1	81,07	46,21	522,45	310,19	-107,00	39,64
2	86,30	44,00	544,35	311,13	-106,13	43,15
3	85,67	44,40	548,98	311,13	-100,25	41,77
4	88,48	43,79	537,25	311,13	-107,63	43,64
5	84,25	43,41	565,64	311,13	-106,31	44,43
6	82,44	44,41	550,86	311,13	-106,13	41,26
7	77,39	43,59	548,66	310,19	-99,75	39,23
8	85,84	42,82	550,05	311,13	-107,88	43,95
9	79,12	43,99	537,97	311,13	-86,31	42,41
10	80,60	43,22	545,02	311,13	-107,88	43,08
11	79,85	43,39	542,79	310,19	-102,06	40,75
12	81,88	42,62	547,68	311,13	-90,81	42,52
13	84,80	43,99	564,33	311,13	-107,88	43,00
14	86,16	42,00	535,00	312,00	-107,88	43,87
15	84,96	42,19	553,71	310,19	-107,88	45,12
16	83,20	42,80	548,82	310,19	-107,00	42,84
17	81,51	42,42	535,78	310,19	-107,19	41,94
18	87,92	42,19	552,11	310,19	-106,31	45,89
19	75,46	43,20	528,25	311,13	-107,88	38,81
20	79,51	43,60	507,27	311,13	-107,88	41,95
21	80,60	42,02	557,29	311,13	-87,88	43,30
22	93,22	44,99	562,11	311,13	-107,88	49,23
23	78,18	42,40	538,62	311,13	-107,00	39,43
24	81,28	42,40	544,29	311,13	-107,63	43,25
25	81,46	42,80	552,60	310,19	-107,00	44,03

**Tabela 1b.** Dados processados por Volta para Predição de Energia

Volta	Temperatura da Chapa de Arrefecimento (oC)	Temperatura do Rotor (oC)	Temperatura do Estator (oC)	Temperatura da Bateria (oC)
1	37,06	34,29	24,16	33,00
2	39,38	36,13	26,04	33,00
3	41,15	38,17	27,94	33,00
4	42,81	39,76	29,16	33,00
5	44,20	41,52	30,62	33,00
6	45,18	42,65	31,54	33,00
7	45,50	44,07	32,39	33,00
8	46,74	45,21	33,37	33,00
9	47,56	46,15	34,15	33,00
10	48,31	47,21	34,74	33,00
11	49,02	48,16	35,60	33,00
12	49,36	48,96	36,02	33,00
13	50,17	49,83	36,28	33,99
14	50,63	50,65	36,98	34,00
15	51,29	51,44	37,70	34,21
16	51,58	52,13	37,94	35,00
17	52,09	52,95	38,51	35,00
18	52,39	53,54	38,75	35,00
19	52,58	54,19	39,21	35,00
20	52,83	54,58	39,11	35,00
21	53,15	55,13	39,73	35,13
22	53,89	55,52	40,07	35,76
23	53,96	56,42	40,55	36,00
24	54,04	56,85	40,55	36,00
25	54,18	57,29	40,86	36,35

Então, a partir do PCA, obteve-se os resultados apresentados na Tabela 2. Pelo critério de Joliffe, foram selecionados aqueles atributos que possuíam autovalor  $\lambda_i > 0.7$ , por apresentarem maior correlação ao consumo de energia. Estes seguem destacados em cinza.

**Tabela 2.** Dados da Análise de Componentes Principais (PCA)

	V1	V2	V3	V4	V5	V6	V7	V8	V9
Tempo de Volta (s)	-0,07	-0,02	0,02	0,74	0,23	0,63	-0,02	0,01	0,03
Velocidade Máxima (rpm)	-0,07	-0,01	-0,18	0,02	-0,01	0,00	0,01	0,01	0,00
Torque Máximo (Nm)	0,00	0,00	0,04	-0,13	-0,09	0,22	0,96	0,02	0,06
Torque Regenerativo Máximo (Nm)	-0,06	0,95	0,29	0,04	-0,04	-0,02	-0,01	-0,01	-0,01
Demanda Média de Pedal (%)	0,50	0,02	0,04	-0,20	-0,04	0,34	-0,06	-0,56	-0,52
Temperatura da Chapa de Arrefecimento (oC)	0,69	0,02	0,04	0,26	-0,02	-0,22	0,08	0,59	-0,21
Temperatura do Rotor (oC)	0,49	0,03	0,02	-0,14	-0,04	0,20	-0,12	-0,13	0,82
Temperatura do Estator (oC)	0,10	-0,03	0,01	0,53	-0,07	-0,58	0,20	-0,56	0,11
Temperatura da Bateria (oC)	0,04	-0,10	0,46	-0,14	0,85	-0,14	0,08	-0,04	0,02
Autovalores	1,20	1,00	0,70	0,40	0,20	0,20	0,00	0,00	0,00

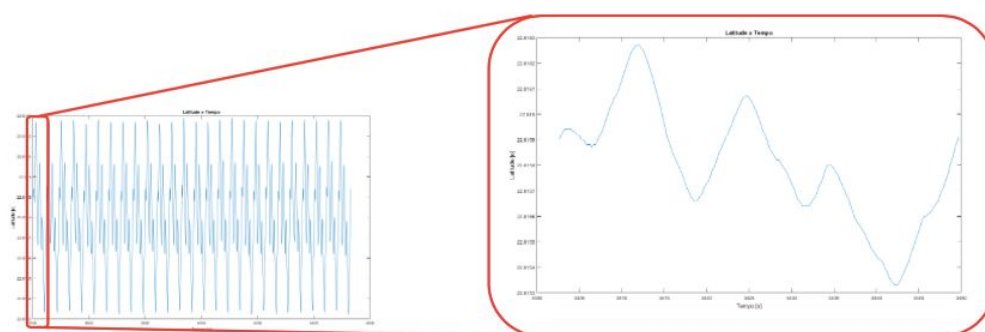
Sabendo que a potência mecânica pode ser descrita pela equação 1, que a energia gasta pode ser descrita pela equação 2, o resultado apresentado pelo PCA se torna claro: a velocidade e o torque estão diretamente relacionados com a potência, que é determinante para o gasto de energia. Além disso, uma vez que a energia é uma integral no tempo, se torna clara também a influência do tempo de volta.

$$E = \int P \cdot dt \quad (1)$$

$$P_{mec} = T \cdot \omega \quad (2)$$

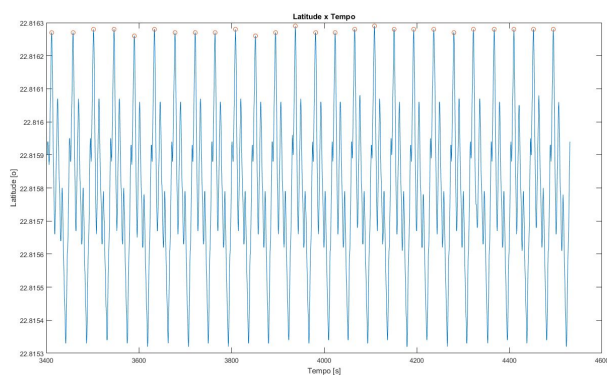
### Determinação de Voltas Dadas

Inicialmente, foi implementado um algoritmo para Correlação Cruzada: o padrão dos dados da primeira volta foram procurados no restante do sinal para determinar em qual instante de tempo cada volta se inicia ou termina. Uma ilustração desta etapa segue apresentada na Figura 6.

**Figura 6.** Procura por Padrão a partir de Correlação Cruzada

Contudo, uma vez que este apenas aplica atrasos, mas não distorce sinais no tempo, foi necessário um *fine tuning*, baseado na ideia de que o piloto não mantém o mesmo tempo por volta. Para isso, foram então encontrados os picos, apresentados na Figura 7, e o tempo inicial e final das voltas foram então ajustados para um bom casamento com cada um dos picos.





**Figura 7.** *Fine tuning* na Busca por Padrão do Sinal

Por último, para a fase de treinamento, os dados foram separados para Treinamento e Validação. Esta escolha foi feita baseando-se no gasto final energético de cada um dos percursos, conforme apresentado na Tabela 3. A fim de verificar se houve um overfitting na Regressão, foi escolhido um percurso bastante distinto da maioria dos outros. O resultado obtido segue apresentado na Tabela 4.

**Tabela 3.** Separação dos Dados para Treinamento e Validação baseado no gasto energético

	Energia Gasta	
<b>Caso 1</b>	1.22 kW	Treinamento
<b>Caso 2</b>	3.57 kW	
<b>Caso 3</b>	3.17 kW	
<b>Caso 4</b>	4.43 kW	
<b>Caso 5</b>	1.64 kW	Validação

**Tabela 4.** Resultado da Validação da regressão feita

	Energia Predita	Energia Gasta	Erro de Predição
<b>Caso 5</b>	1.51kW	1.64kW	8.3%

## CONCLUSÕES

Considerando que, para esta aplicação, um erro de 8% pode ser determinante para o motorista ser colocado em uma situação de risco (por exemplo, se ficar parado no meio de uma estrada deserta) ou um atraso significativo para um compromisso, uma vez que a carga de baterias é bastante lenta, o resultado obtido não é satisfatório.

Sendo assim, são sugeridas novas implementações, tais como outros algoritmos de predição e algoritmos de predição por volta em vez de percurso completo; uma base de dados maior se vê também como necessária para o bom desenvolvimento deste projeto. Por fim, sugere-se como trabalho futuro utilizar modelos de veículos em simulações *In-The-Loop* para validação do modelo desenvolvido.

## **REFERÊNCIAS**

**BRAUN M.R., ALTAN H., BECK S.B.M.**, Using regression analysis to predict the future energy consumption of a supermarket in the UK, 2014.

**MOTTAHEDIA M., MOHAMMADPOUR A., AMIRIB S.S., RILEYB D. ASADIB S.**, Multi-linear Regression Models to Predict the Annual Energy Consumption of an Office Building with Different Shapes, 2015.

**ALVES, R. S. S. V.**, Application of Neural Networks' Models to Predict Energy Consumption, 2016.

**PANDEY, R. K.**, Data Quality in Data warehouse: problems and solution, 2014.

**KIMBALL R., ROSS M.** The Data Warehouse Toolkit, 2nd edition, 2002.

**SOUCY, P., MINEAU, G. W.**, A simple KNN algorithm for text categorization, 2001.

**WONG J. W. H., DURANTE C., CARTWRIGHT H. M.** Application of Fast Fourier Transform Cross-Correlation for the Alignment of Large Chromatographic and Spectral Datasets, 2005.