VYTAUTO DIDŽIOJO UNIVERSITETAS
Informatikos fakultetas

APPLICATION OF MACHINE LEARNING FOR IMPROVING CONTEMPORARY ETA FORECASTING

Transport management systems, as well as tracking and monitoring systems, are used by participants in a supply chain to improve transparency. It is becoming increasingly common for vehicles to be equipped with improved microprocessors, larger memory capacity, and real-time operating systems. As a result of the newly installed technological platforms, the operating system has the ability to handle more advanced applications such as model-based process control, artificial intelligence, and comprehensive computations. The purpose of this study is to identify methods that are best suited to solve the problem of estimating the arrival time (ETA) by ranking the drivers based on drivers' behaviour data and estimating deviations from planned arrival time using different machine learning methods.

RANKING DRIVERS

Data:

- from August 21, 2020, to January 1, 2022
- 398 observations
- attributes from a vehicle monitoring system:
  1) Free-rolling distance,
  2) Engine overloaded distance,
  3) Highest gear distance,
  4) Excess idling,
  5) Overspeeding time,
  6) Extreme braking events,
  7) Harsh braking events.

Selection of attribute weights:

Two sets of weights were used:

1) weights with equal importance for all attributes
2) weights in accordance with the evaluation by an expert.

Attribute	Score	Weight w_k	min/max
Free rolling distance	5	0.09	max
Engine overloaded distance	10	0.19	min
Highest gear distance	7	0.13	max
Excess idling	10	0.13	min
Overspeeding time	10	0.13	min
Extreme braking events	8	0.15	min
Harsh braking events	4	0.07	min

Results:

	Score									
	10	9	8	7	6	5	4	3	2	1
TOPSIS W_1	325	62	6	4	0	0	1	0	0	0
TOPSIS W_2	316	71	7	3	0	0	1	0	0	0
VIKOR W_1	156	107	67	28	16	6	9	3	5	1
VIKOR W_2	148	120	63	31	12	8	6	5	4	1

Considering the results of these ranking methods, the method with the most logical ranking of the drivers was confirmed by an expert to be using the VIKOR method.

FORECASTING THE ETA

Data:

- from August 21, 2020 to January 24, 2022
- 1758 observations
- logistics data variables:
  1) Driver's score
  2) Tour beginning country
  3) Tour ending country
  4) Furthest country
  5) Tour beginning day
  6) Tour beginning month
  7) Number of intermediate stops
  8) Vehicle height
  9) Vehicle width
  10) Vehicle length
  11) Vehicle weight
  12) Hours of service breaks
  13) Planned distance

Predictor:

$$\Delta t_i = T_i - t_i$$

where

Δt_i - the difference between the planned and factual delivery time,
T_i - the factual time when the i^{th} cargo will be delivered,
t_i - the planned time of delivery for the i^{th} cargo,
$i = 1,2,...,n.$

Analysis and Results:

- Normalization. Quantitative variables were normalized using min-max normalization.
- Transformation. The categorical variables were transformed into dummy variables.
- Train-test split. 75% of the dataset was assigned to the training sample, 25% to the test sample.
- Selection of optimal parameters. Cross-validation (k = 10) was used.

Model	R^2_{adj}	RMSE	MAE
Decision tree	0.6666	1391.18	792.56
Random forest	0.7757	1120.15	683.94
XGBoost	0.7427	1210.98	777.18
SVM	0.6726	1408.97	867.82
KNN	0.2465	2105.53	1208.22
Ensemble model	0.7795	1118.28	683.47

A possibility to improve the models by forming an ensemble model

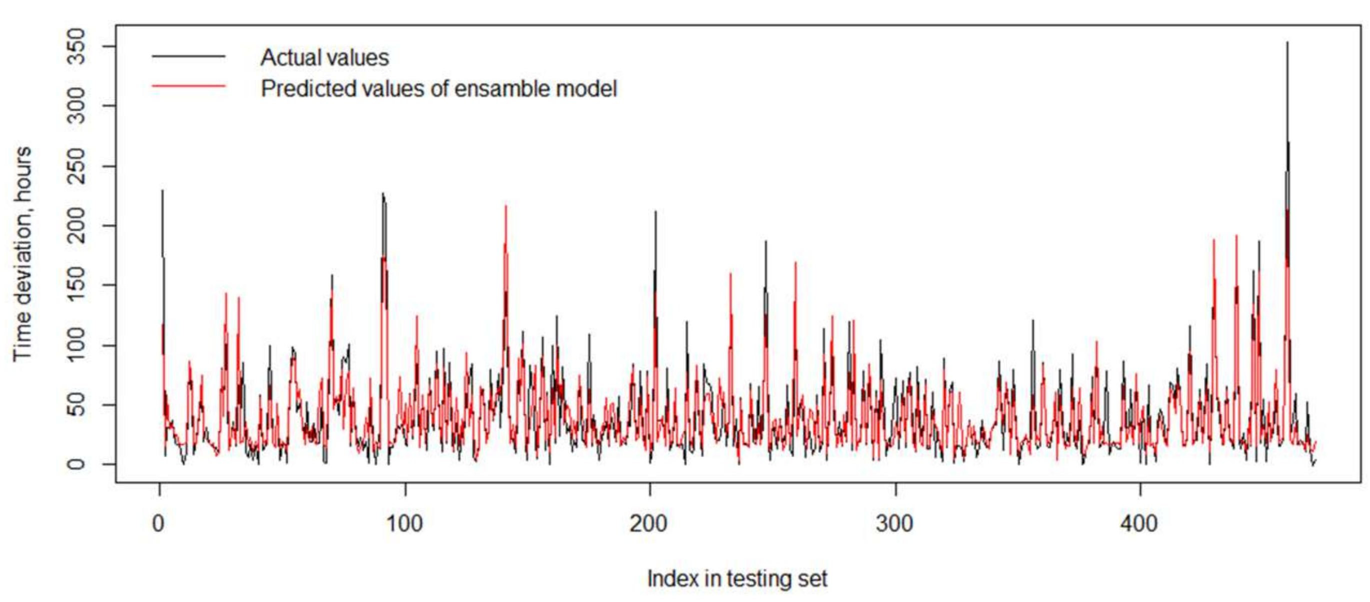
$$\hat{y}_i = (1 - c_2) \times \hat{y}_{i\,rf} + c_2 \times \hat{y}_{i\,svm},$$

where

$\hat{y}_i$ - the predicted value of the i^{th} observation for the model ensemble,
$\hat{y}_{i\,rf}$ - the predicted values of the ith observation of the random forest model,
$\hat{y}_{i\,svm}$ – for the SVM model.

The experiment resulted in the most accurate results ($R^2_{adj} = 0.7795$) when c_2 was equal to 0.2

$$\hat{y}_i = 0.8\,\hat{y}_{i\,rf} + 0.2\,\hat{y}_{i\,svm}.$$



AUTHORS:

Gabrielė Kasputytė
gabriele.kasputyte@vdu.lt

Arnas Matusevičius
arnas.matusevicius@vdu.lt

Tomas Krilavičius
tomas.krilavicius@vdu.lt

Ruptela Transport Telematics

CARD
CENTRE FOR APPLIED RESEARCH AND DEVELOPMENT

ACKNOWLEDGMENTS

We thank Tadas Motiejūnas, Neringa Andrulytė, Darius Ežerskis, Viktorija Naudužaitė, Donatas Šumyla and UAB Ruptela for cooperation and useful insights.