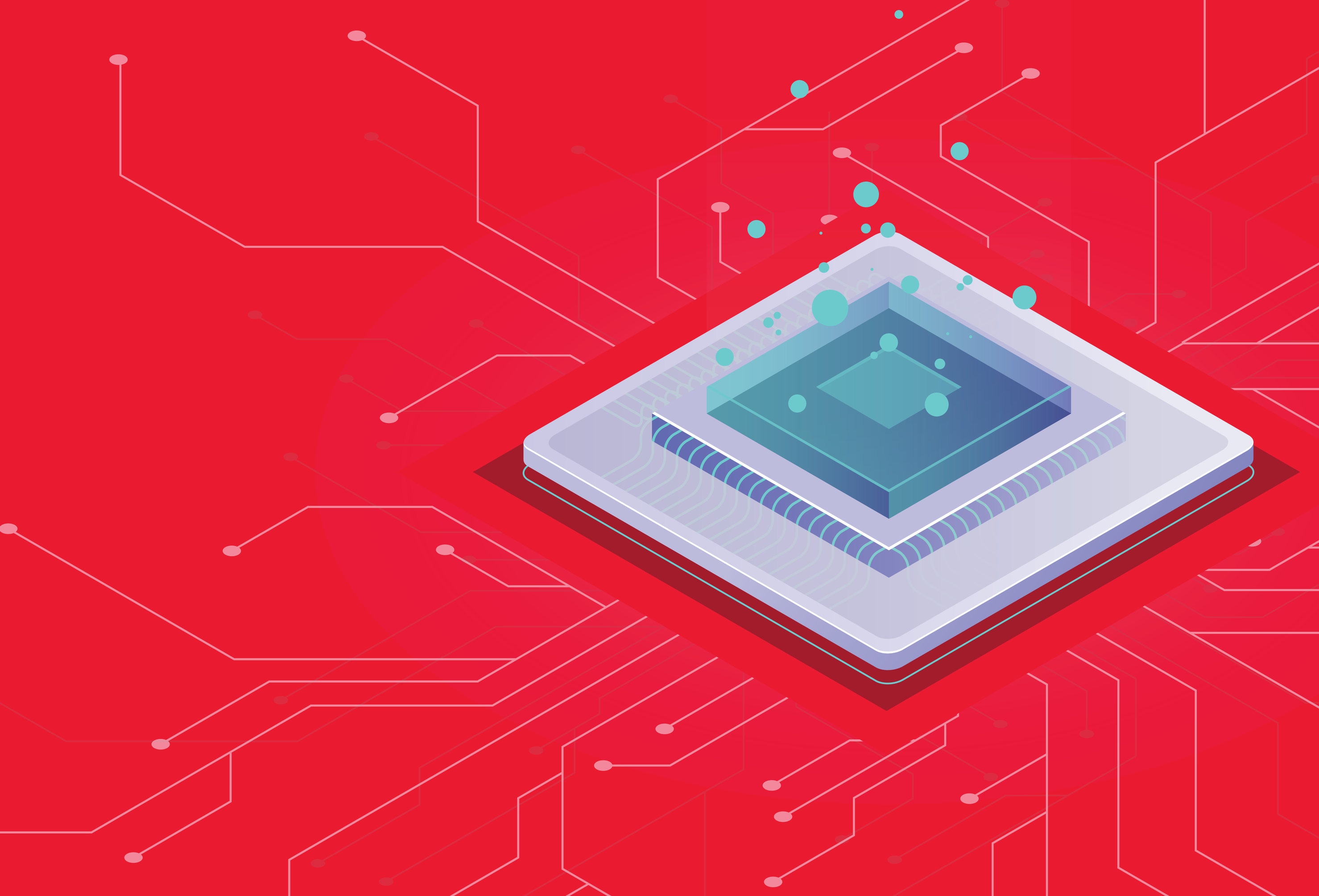
SKIP THE DISHES

Predicting Food Preparation Time

Joshua Hamilton Sehaj Pal Singh Jaret André



PROBLEM STATEMENT

SkipTheDishes drivers make dozens of food-deliveries every day. They are sent requests by customers to pick up food from specific restaurants and deliver that food to the customer's requested location. To effectively coordinate this complex process between Drivers, Restaurants, and Customers, we will take a Data-Driven approach. In this project, we aim to predict, ahead of time, how long a restaurant will take to prepare the ordered food so that the Driver will be able to pick up the food on time and deliver it to the customer as quickly as possible.

DATA PREPERATION

We added five new categories to the menu data set to fill in it's missing values- 'baking,' 'cooking,' 'red_meat,' 'seafood,' and 'white_meat.' We also denoted whether each menu item qualified as a 'starter,' a 'main' or a 'dessert.' In our training and test data sets, we then recorded the quantity of each category ordered for each row. Rather than paying attention to the order in which each of the requested food-items appear in each row, we decided to pay attention to the items that were requested, strictly - and how much. And lastly, we Cyclically-Encoded our time variables so that our model would be less likely to overfit and also so that it could more easily recognize times that fall close together.

a Original Date Time

	datetime
0	2019-09-03 10:56:39
1	2019-09-04 21:59:57
2	2019-09-04 12:50:51
3	2019-09-05 20:48:40
4	2019-09-05 14:06:43

b Separated Date Time

	day	hour	minute
0	3	10	56
1	4	21	59
2	4	12	50
3	5	20	48
4	5	14	6

$$\text{Hour} \longrightarrow \begin{cases} \sin\left(\frac{\text{Hour}}{24}2\pi\right) \\ \cos\left(\frac{\text{Hour}}{24}2\pi\right) \end{cases}$$

c Cyclic-Encoded Date Time

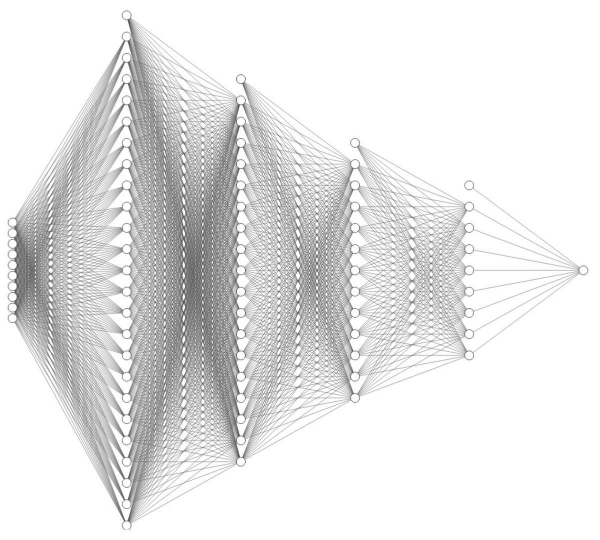
	(1.0,)	(2.0,)	(3.0,)	(4.0,)	(5.0,)	(6.0,)	(7.0,)	hour_SIN	hour_COS	minute_SIN	minute_COS
0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.398401	-0.917211	-3.140767e-01	0.949398
1	0.0	0.0	0.0	1.0	0.0	0.0	0.0	-0.519584	0.854419	-2.449294e-16	1.000000
2	0.0	0.0	0.0	1.0	0.0	0.0	0.0	-0.136167	-0.990686	-8.183028e-01	0.574787
3	0.0	0.0	0.0	0.0	1.0	0.0	0.0	-0.730836	0.682553	-9.213120e-01	0.388824
4	0.0	0.0	0.0	0.0	1.0	0.0	0.0	-0.631088	-0.775711	5.963674e-01	0.802712

METHOD

After cleaning our data (imputing missing values in the dataset as well as adding new features), we decided to use a Feed-Forward Neural Network as our Model. After tuning various hyperparameters, we settled on a four-layer NN with 109 input neurons, 672 interior neurons and a single output neuron. We integrated the Batch-Normalization technique into our NN, and used the ReLU activation function with each neuron.

$$\text{ReLU}: f(x) = \max\{0, x\}$$

FeedForward Neural Network



RESULTS

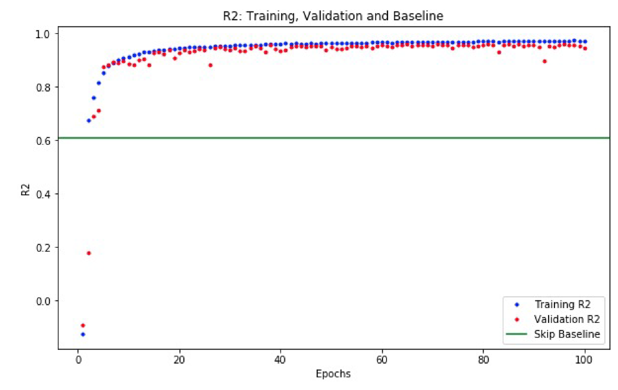
After finding our optimal hyperparameters, splitting our data into a training and a validation set (80/20 split), we obtained the following metrics on our validation set:

R^2 = 0.9840358057780032
MAE = 1.091128219485283
RMSE = 1.433299268900221

In contrast, the following baseline scores were provided to us by the *SkipTheDishes* team at the beginning of the competition:

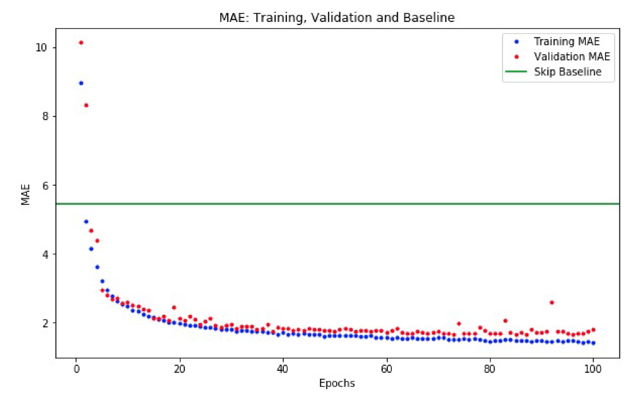
R^2 = 0.6100377
MAE = 5.448831
RMSE = 6.978090

a R^2



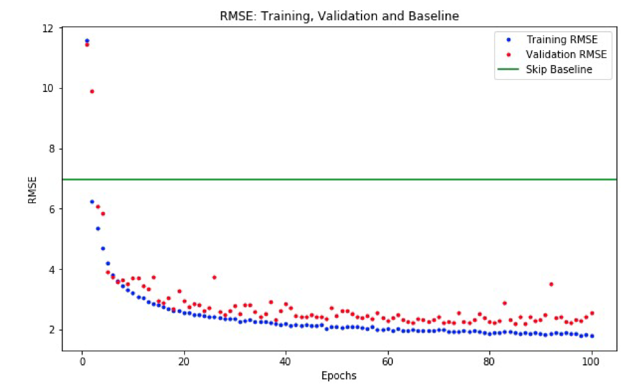
$$R^2 = 1 - \frac{SSE}{SST}$$

b MAE



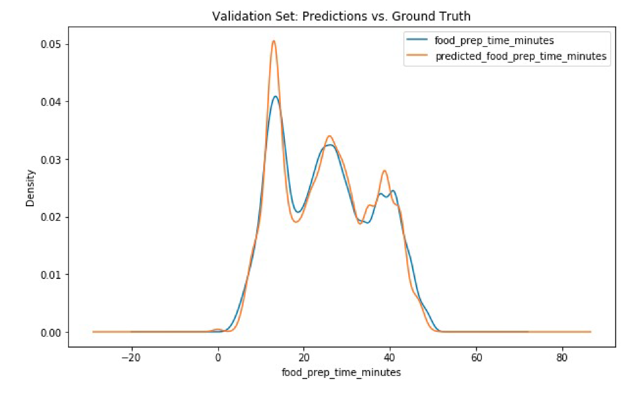
$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$

c RMSE



$$RMSE = \sqrt{MSE}$$

d Predictions Against Validation Set



faculty of SCIENCE
discover the unknown + invent the future

CONCLUSION

Our observations indicate that our model's performance has significantly increased the scores of our metrics from that of the original baseline scores. It would appear that a Feedforward Neural-Network Regression model, with the appropriate data-cleaning procedure, can help us accurately predict the food preparation time of orders. However, since Neural Networks are highly scalable models, it remains to be seen how this model would do if it were exposed to more data.

ACKNOWLEDGEMENTS

We would like to sincerely thank the Nexus Data Science Institute for hosting this fantastic event and SkipTheDishes for providing us with a fun and challenging data science problem. We want to extend our gratitude to Dr. Carson Leung for allowing us to work and collaborate in his Data Mining Laboratory, including the Faculty of Science for providing us with additional resources.