Faculty of Computer Science

Master of Data Science
Fall 2020

Moscow
2022

Choice of Metrics in ML-Based Forecasting of Ingredients in Retail

Student: Vorobeva Tatiana

Supervisors:

- Polina Polunina, visiting lecturer, Big Data and Information Retrieval Department, Faculty of Computer Science; Specialist McKinsey and Co
- Sarkis Grigoryan, Head of Competence center for Artificial Intelligence at Digital Economy Development Fund

Agenda

1. Introduction
    - Topic relevance
    - Business problem statement
    - Research objectives
2. Metrics classification and calculation
3. Experiments description
    - Data description & preprocessing
    - Model choice
    - Metrics exploration
4. Conclusion

Introduction

Topic relevance

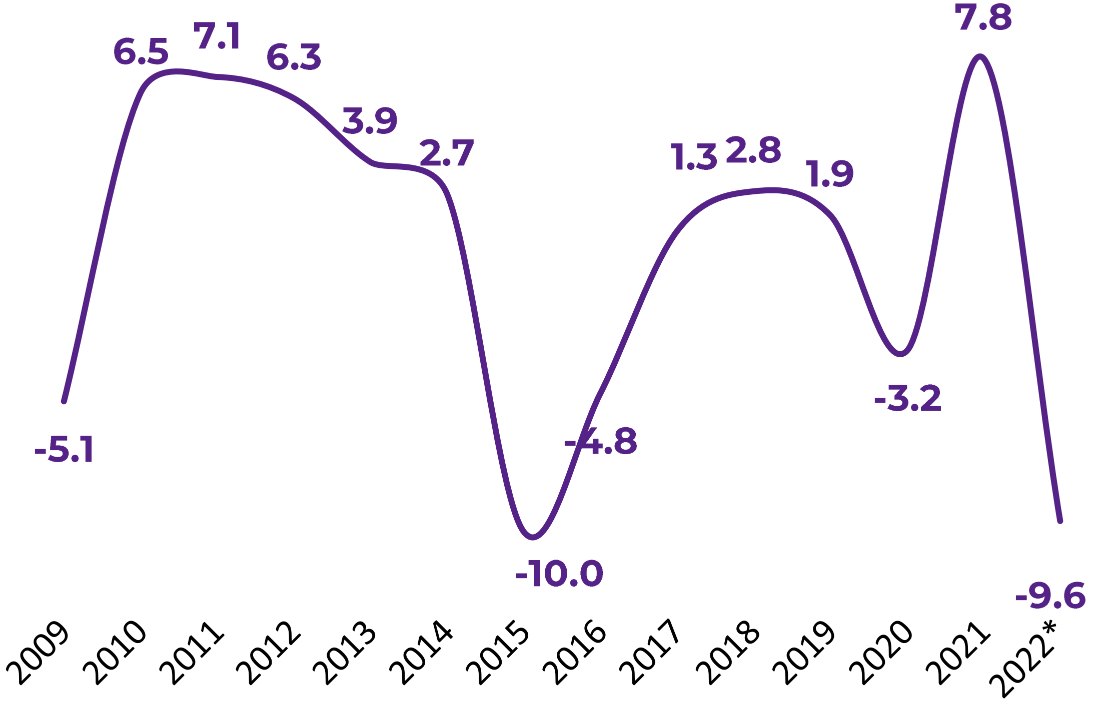
Recent changes in retail / QSR* landscape

Increased importance of operational efficiency due to turbulent economic situation

Need to track model quality & adjust accordingly to changing factors

Limited number of case studies on forecast accuracy metrics

Retail turnover, %



*QSR – quick service restaurants
Source: Rosstat, Ministry of Economic development forecast

Introduction

Business problem statement (1/2)

Data for this research was provided by “Teremok” company

- Chain of **quick service restaurants**, specializing on Russian cuisine
- **165 restaurants** located in Moscow, Saint-Petersburg and Krasnodar
- Employs **over 5000 people**

Source: Teremok official website https://teremok.ru/

Introduction

Business problem statement (2/2)

Forecasting model structure

INPUT

Data for the last 14 days:

- Stock
- Consumption
- Write-offs
- Sales stop

Model

OUTPUT

Order of ingredients for next day:

- Number of boxes of each ingredient

Key features of procurement process:

1. Orders are placed by each restaurant **for next day**
2. Ingredients are ordered **by boxes**
3. Lost sales driven by **out-of-stocks** lead to **higher profit losses than** potential **write-offs** of excessive ingredients

Introduction
Research objectives

Research objectives:

1. **Identify** the **most applicable** ML-based **forecasting model** based on current business process and specifics of data
2. Propose a **set of metrics** which would enable **more efficient model quality** and business **risk monitoring**

Methods:

Overview of **related work and theoretical concepts**, mathematical **modelling**, **analysis of formal outcomes and** their translation into **recommendations** for business

Methodology

Metrics classifications

Makridakis and Hibon (1995)

1° dimention:

- Absolute
- Relative to a base or other method
- Relative to the size of errors

2° dimention:

- Single method
- More than one method
- In comparison to some benchmark

Hyndman (2006)

- Scale-dependent (e.g. MAE, GMAE);
- Percentage-error (e.g. MAPE);
- Relative-error (e.g. MdRAE, GMRAE);
- Scale-free error (e.g. MASE)

Botchkarev (2019)

- Primary
  - Plus typology across 3 key components
- Extended
- Composite
- Hybrid

Source: based on Hyndman (2006) [4], Botchkarev (2019) [5]

Methodology

Key metrics overview

	MAE	MdAE	MSE	RMSE	BIAS	MAPE	MdAPE	WAPE
Type	Absolute					Percentage-based		
Calculation	$\frac{1}{n}\sum_{t=1}^{n}\lvert y_t-\hat{y}_t\rvert$	$\underset{t=1\ldots n}{\mathrm{median}}(\lvert y_t-\hat{y}_t\rvert)$	$\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2$	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}$	$\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)$	$\frac{1}{n}\sum_{t=1}^{n}\frac{\lvert y_t-\hat{y}_t\rvert}{y_t}$	$\underset{t=1\ldots n}{\mathrm{median}}\left(\frac{\lvert y_t-\hat{y}_t\rvert}{y_t}\right)$	$\frac{\sum_{t=1}^{n}\lvert y_t-\hat{y}_t\rvert}{\sum_{t=1}^{n}\lvert y_t\rvert}$
Aggregation	Mean	Median	Mean	Mean	N/A	Mean	Median	Weighted Mean
Comparability across scales	-	-	-	-	+/-	+	+	+
Resistance to outliers	-	+	-	-	-	-	+	+
Handling zero observations	+	+	+	+	+	-	-	+

Source: compiled by author based on [4], [5]

Experiments

Data description

Restaurants

- 9 restaurants
- 3 groups (with high / medium / low turnover)
- By 3 restaurants in each group

Time period

- Daily data
- 2 years

Features

- Ingredient id
- Stock
- Consumption
- Write-offs
- Sales stop
- Packaging coefficient

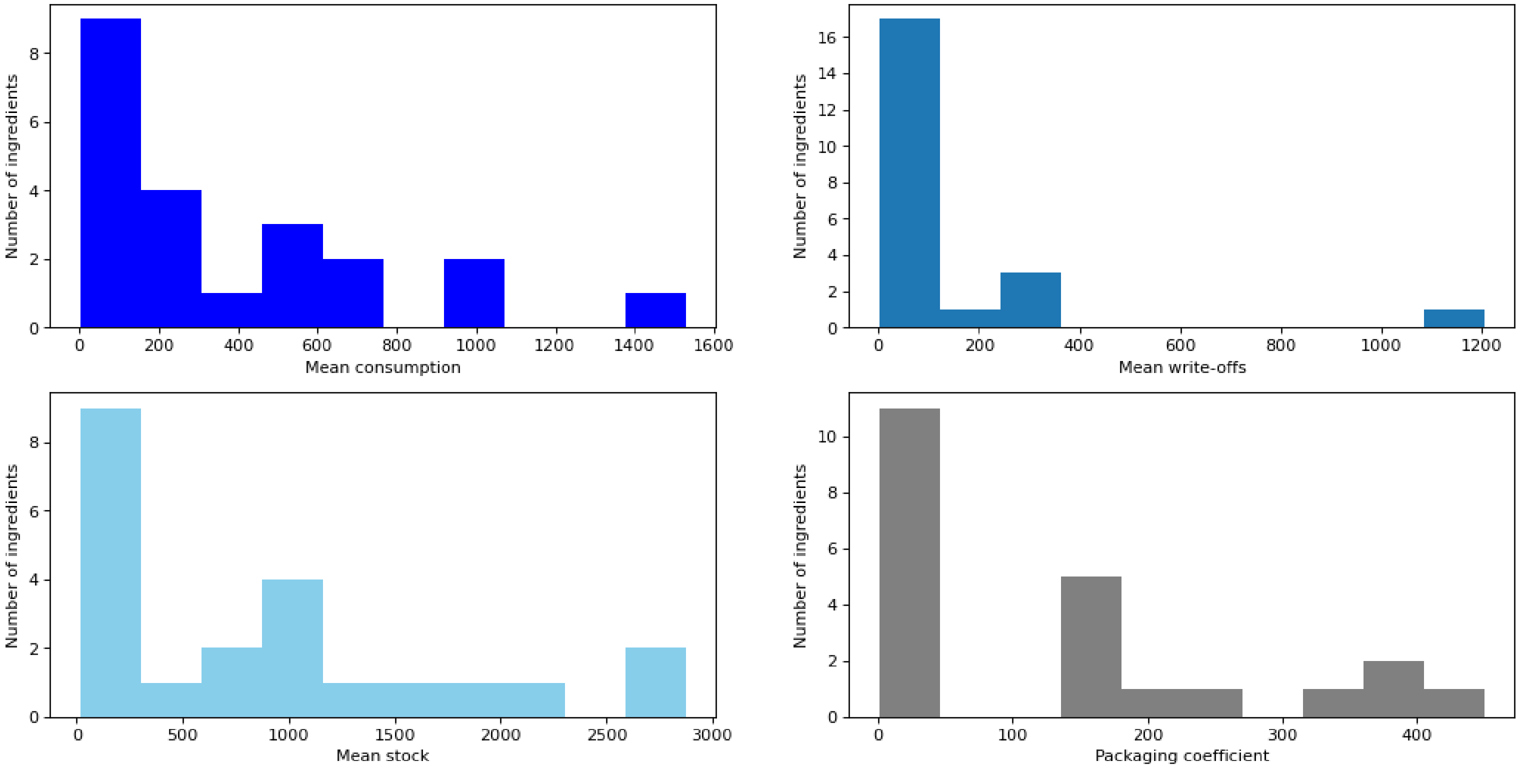
Experiments

Data description

Explorative analysis of the dataset showed the following specifics:

1. **Presence of outliers**
2. Difference in KPIs by groups of restaurants

Features distribution



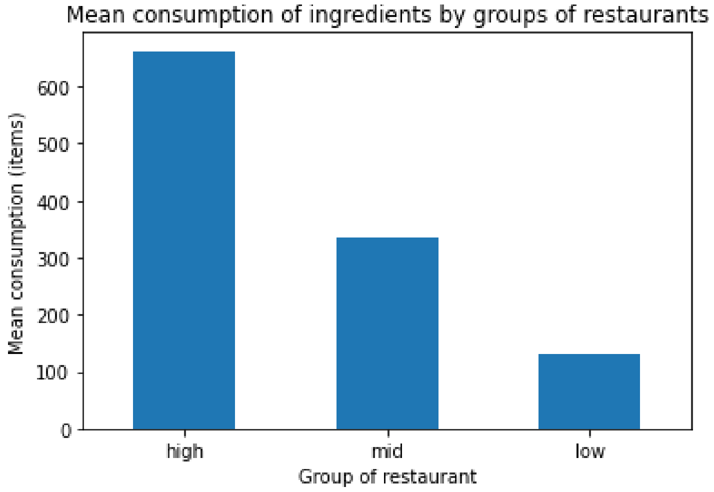
Experiments

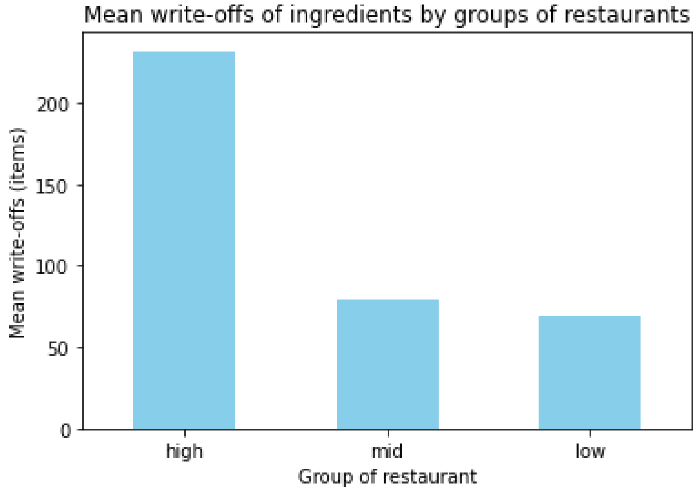
Data description

Explorative analysis of the dataset showed the following specifics:

1. Presence of outliers
2. **Difference in KPIs by groups of restaurants**

Mean consumption and wrote-off by groups of restaurants





Experiments

Data preprocessing and target calculation (1/2)

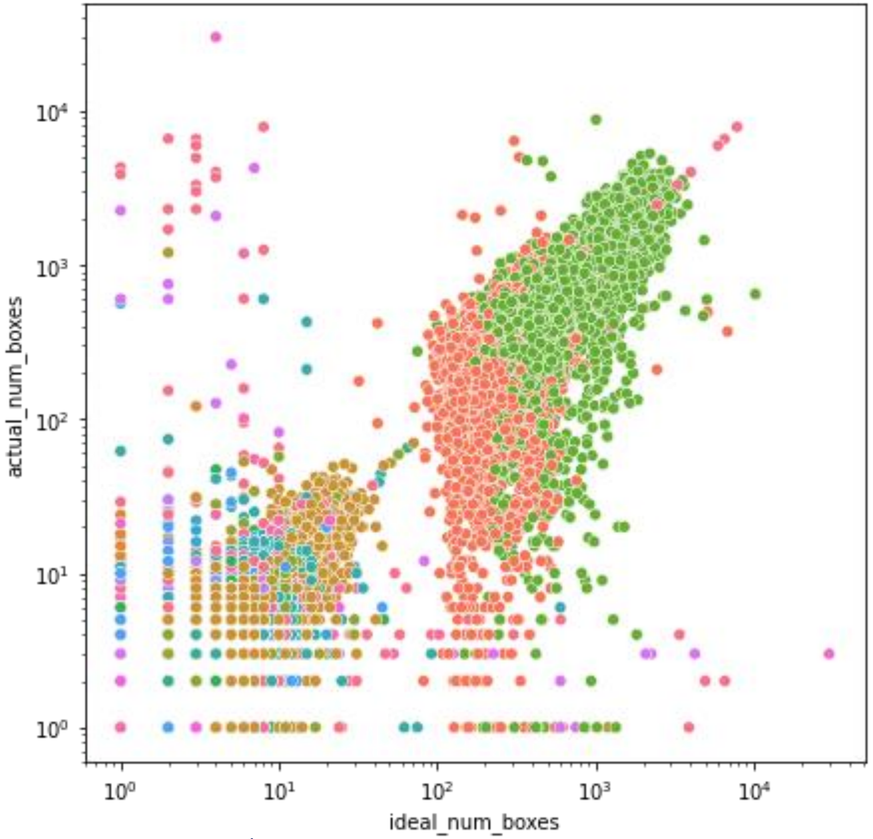
To build a model we preprocessed given dataset:

1. **Observations** were grouped into 14-day periods
2. Actual **number of consumed boxes** of ingredients was calculated
3. **Categorical feature** (item_id) was encoded (OHE)

4. **Target** was calculated **as ideal number of boxes** to order on day D

Ideal order on day D allows to have **stock of ingredient** at the end of the day **equal to half the consumption by 12:00 on day D+1**

Actual vs 'ideal' number of boxes, %



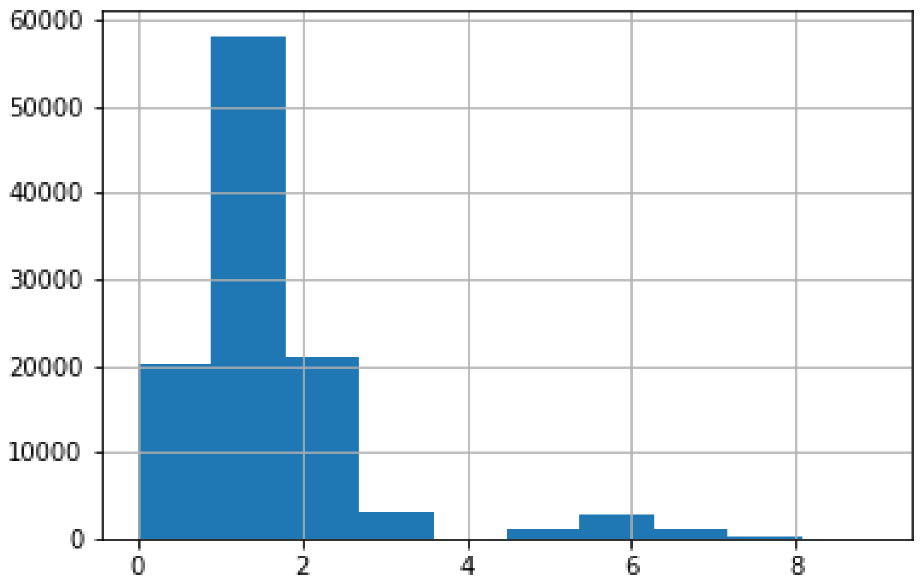
Experiments

Data preprocessing and target calculation (2/2)

- For modelling: logarithm from target as input
- For forecasting: inverse transformation

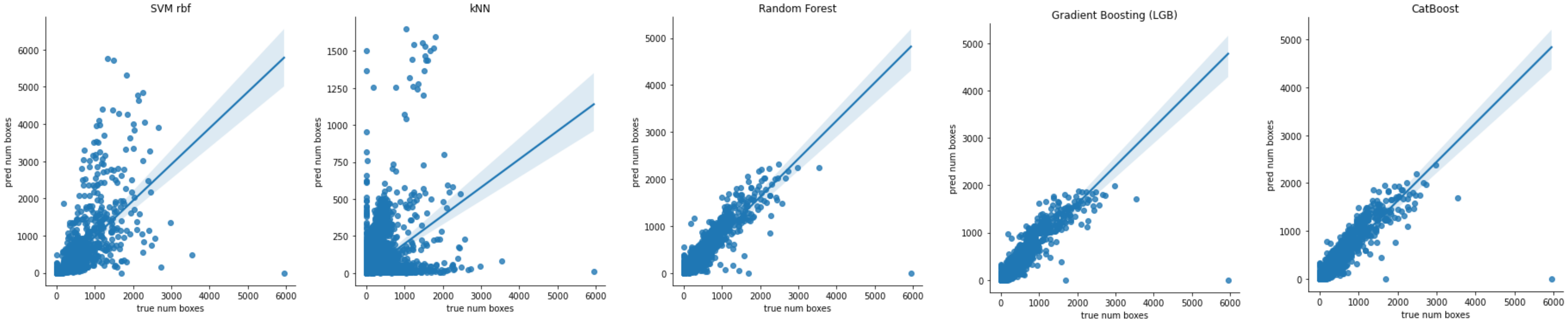
- A ‘tail’ in target distribution is observed, which mostly consists of items-outliers

Distribution of target value, %



Experiments

Choice of model

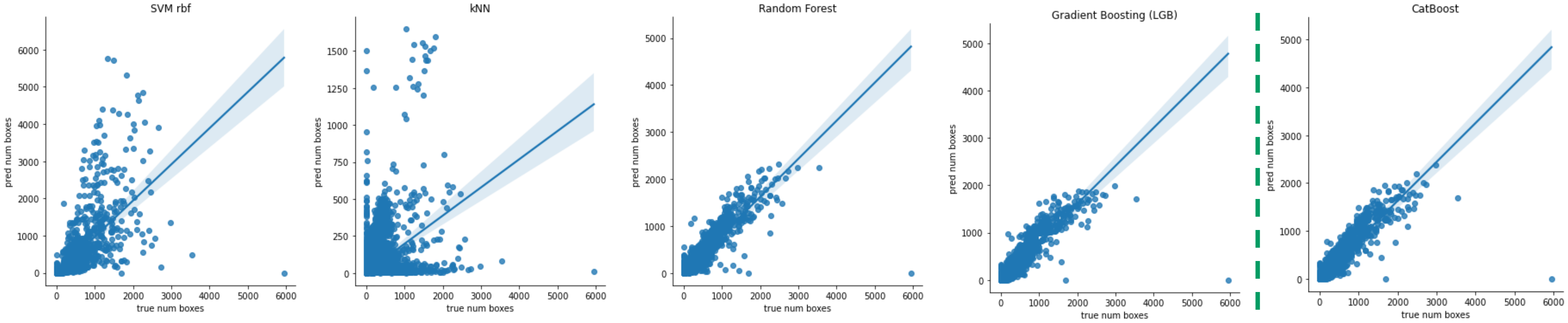


	SVM (RBF)	KNN	Random Forest	LightGBM	CatBoost
R2	0.29	0.22	0.86	0.87	0.87
Spearman	0.39	0.70	0.89	0.89	0.90
RMSE	122.66	129.09	53.64	53.34	52.34
MAPE	94.17	94.43	46.37	44.12	42.07

Experiments

Choice of model

Best result

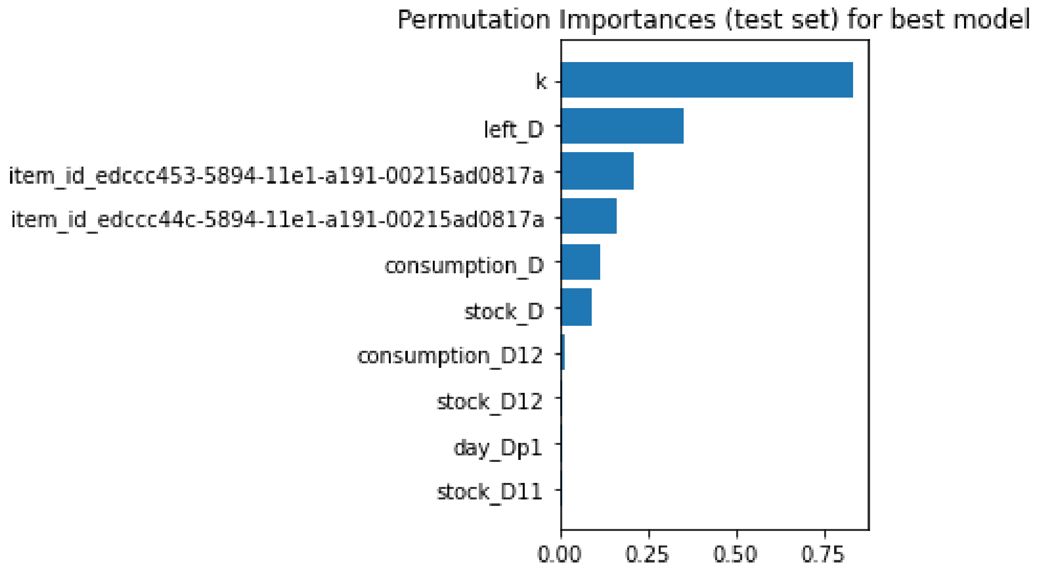


	SVM (RBF)	KNN	Random Forest	LightGBM	CatBoost
R2	0.29	0.22	0.86	0.87	0.87
Spearman	0.39	0.70	0.89	0.89	0.90
RMSE	122.66	129.09	53.64	53.34	52.34
MAPE	94.17	94.43	46.37	44.12	42.07

Experiments

Choice of model

Permutation importance test



Metrics for CatBoost with default vs optimized hyperparameters

metric	Default	Optimized
R2	0.87	0.87
Spearman	0.90	0.90
RMSE	52.34	52.52
MAPE	42.07	40.83

Experiments

Forecast accuracy metrics analysis (1/2)

- **Absolute errors** (MAPE, MAE and BIAS) are very **vulnerable to outliers.**
- **Median errors** (MdAPE, MdAE) and **WAPE** provide more **coherent estimations**.
- **Custom metric** ***Median under-order*** was calculated as: **median** value of all **cases**, when **forecasted number** of boxes to order was **below actually** consumed.

metric	all sample	without outliers
MAPE (mean)	40.8	19.4
MdAPE (median)	14.2	13.7
MAE (mean)	6.83	0.89
MdAE (median)	0.63	0.57
BIAS	2.86	0.18
WAPE	24.5	21.9
Median under-order	-0.03	-0.03

Experiments

Forecast accuracy metrics analysis (2/2)

- **Scale-free (absolute)** errors **tend to be lower** for group of restaurants with low turnover, which might be explained by **overall lower off-takes**.
- **Percent-based** errors like MAPE, MdAPE, WAPE tend to **show more coherent results**.

metric	high	medium	low
MAPE (mean)	37.6	31.9	47.2
MdAPE (median)	13.5	14.5	13.3
MAE (mean)	8.69	7.39	4.78
MdAE (mean)	0.95	0.64	0.37
BIAS	2.6	2.37	2.53
WAPE	18.3	23.9	30.3
Median under-order	-0.06	-0.04	-0.02

Conclusion

Model testing:

- Several models were tested (SVR with different kernels, KNN, Random Forest, Gradient Boosting (LightGBM, CatBoost).
- **CatBoost** showed best results on total sample and sub-samples.

Metrics exploration:

- A **set of standard** metrics was calculated.
- **Custom metric** was proposed.
    - In case of outliers in data **median errors are more robust.**
    - **Absolute** errors have **more physical sense** and can be easily **converted to profit losses** (mind outliers).
    - If **business process** requires **certain conditions** to be met a **custom metric** might be used (mind outliers).

Recommendation: a combination of metrics (e.g. MdAPE, BIAS and Median under-order). It also might be helpful for “Teremok” to approximate forecast errors to lost profits or extra losses.

Thank you!