MBDW 2020

XLAB

Machine Learning based System for Vessel Turnaround Time Prediction

2nd Maritime Big Data Workshop

IEEE International Conference on Mobile Data Management (MDM 2020)

Dejan Štepec	**Tomaž Martinčič**	Fabrice Klein	Daniel Vladušič	Joao Pita Costa
XLAB d.o.o.	XLAB d.o.o.	Port of Bordeaux	XLAB d.o.o.	XLAB d.o.o.

AI enabled port operations

- **Growth** across all the major cargo types
- More than **90%** of the **world’s trade**
- Expected to **annually expand 3.5%** (2019 - 2024)
- Increasing pressure on **ports - turnaround times**
- Digitalization in the ports (**PCS**) and standardized data (**FAL forms**)

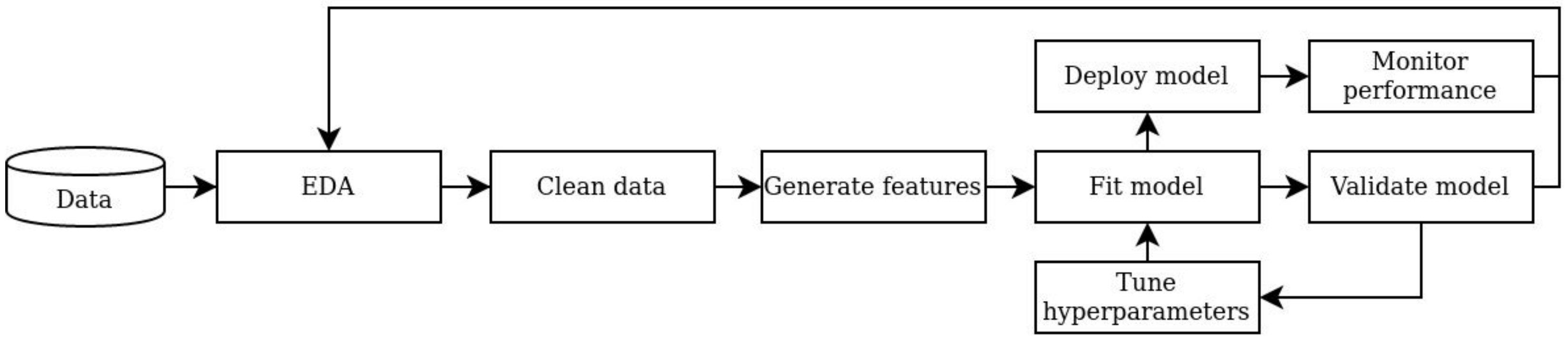
*Vessel turnaround time is one of the main indicators of ports efficiency and trade competitiveness**

*J. Hoffmann and S. Sirimanne, “Review of maritime transport,” in United Nations Conference on Trade and Development, 2019.

Port of Bordeaux (GPMB)

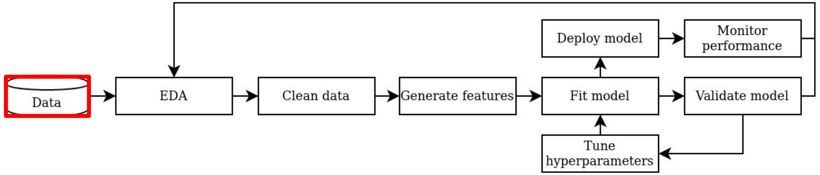
- **Atlantic coast**, just outside of **Bordeaux**
- Focal point of a dense river, sea, air, rail and road traffic network
- Core **TEN-T** port
- 7th of French ports (8 to 9 Mt/year)
- 7 terminals
- Internal PCS - **VIGIEsip**

Turnaround time / ETD workflow



Data

FAL forms and other

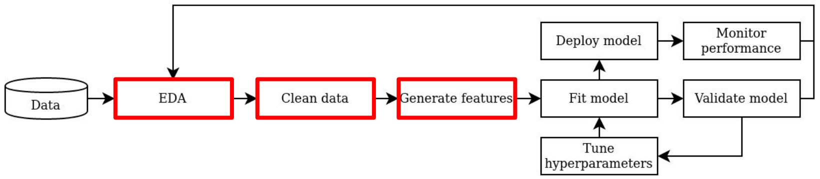


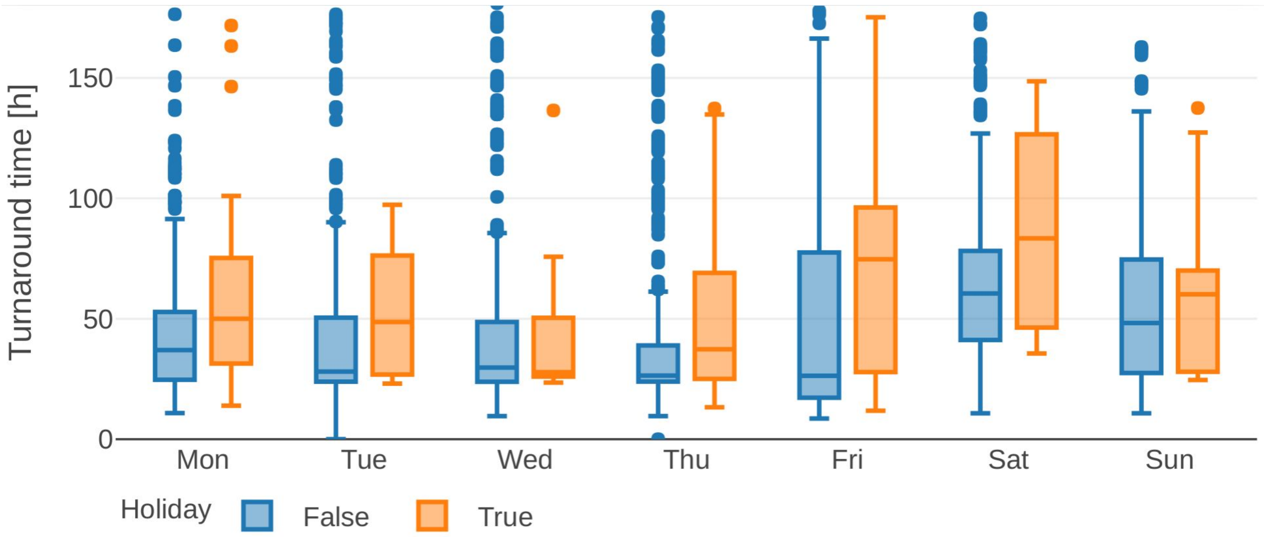
- Captured in VIGIEsip PCS - FAL forms data
- **6055** (1905 unique) vessel arrivals (**2008 - 2018**)
  - Availability to live operational data
- **55** unique loading and **46** unloading cargo types
- Arrival & departure times validated with **AIS data**
- Weather, holidays, AIS and water height

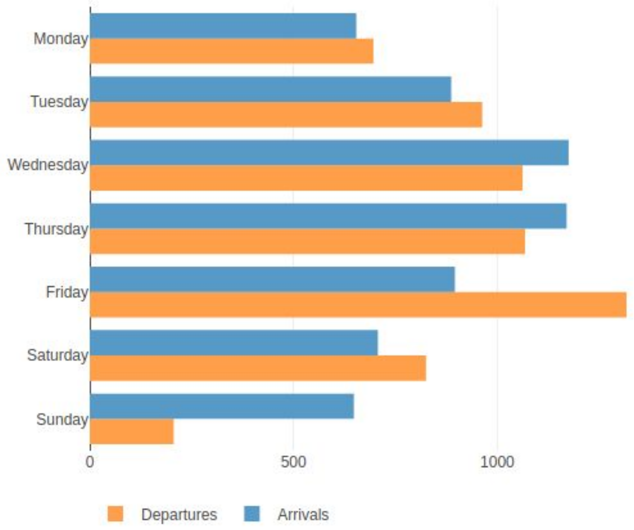
	name	arrival_dock	departure_dock	unloading_berth	unloading_cargo_type	unloading_cargo_fiscal_type	unloading_tonnage	unloading_agent	loading_berth	loading_cargo_type	loading_cargo_fiscal_type	loading_tonnage	loading_agent
0	ISIS	2008-01-02 14:30:00	2008-01-05 03:30:00	433	UREE VRAC	SOL.V	5593	SEAINVEST	433	None	SOL.V	0	SEAINVEST
1	CEC STAR	2008-01-02 23:00:00	2008-01-03 21:20:00	432	CONTENEURS	P.C.	1849	CGM	432	CONTENEURS	P.C.	3084	CGM
2	INGRID JAKOBSEN	2008-01-03 00:25:00	2008-01-05 03:15:00	436	None	LIQ.V	0	SEAINVEST	436	HUILE COLZA	LIQ.V	6000	SEAINVEST
3	HANSEATIC SCOUT	2008-01-03 12:30:00	2008-01-04 13:35:00	449	None	SOL.V	0	SURSOL	449	TOURNESOL VRAC	SOL.V	2612	SURSOL
4	JAC	2008-01-04 15:05:00	2008-01-08 18:20:00	413	ALCOOLS VRAC	LIQ.V	1572	WORMS SM	413	None	LIQ.V	0	WORMS SM

EDA

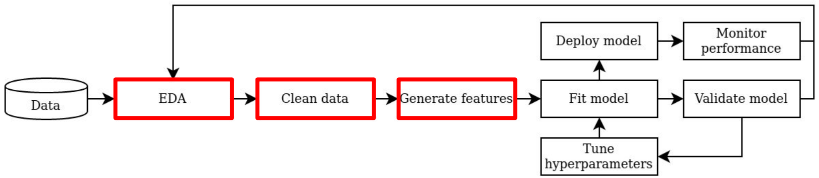
Day of arrival

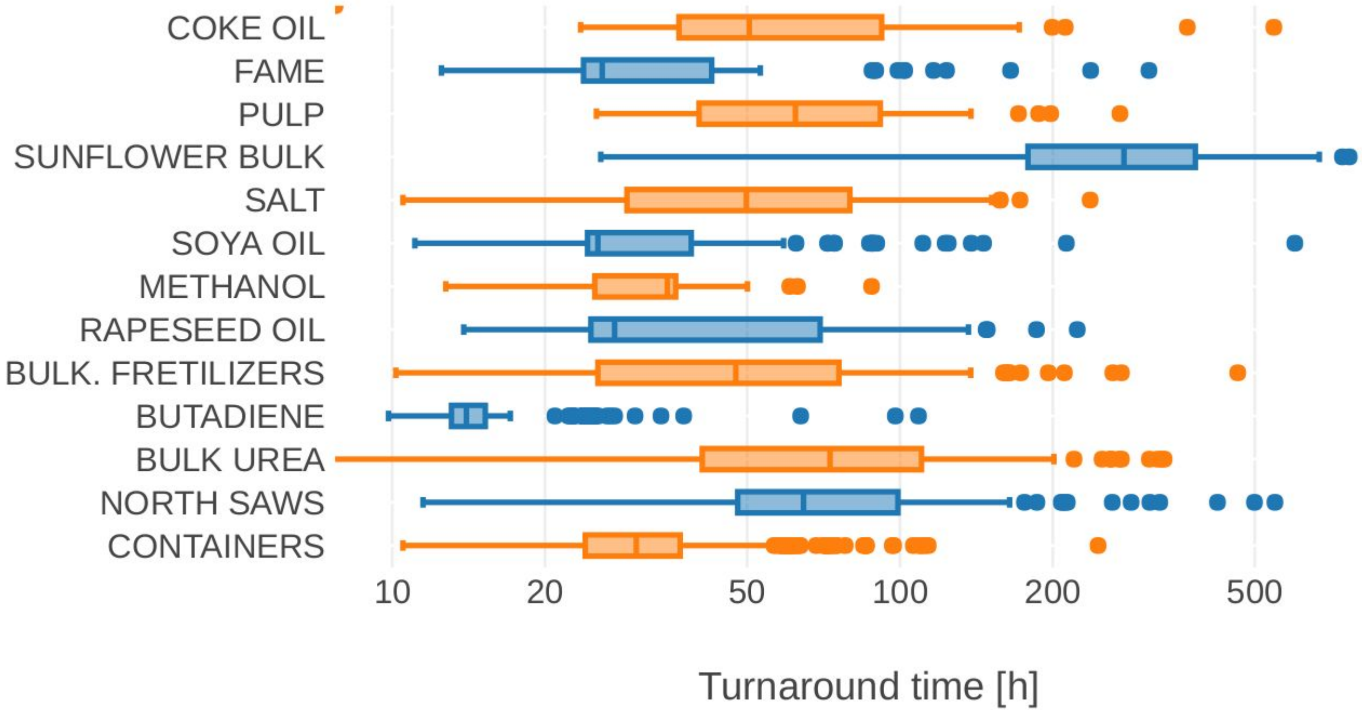






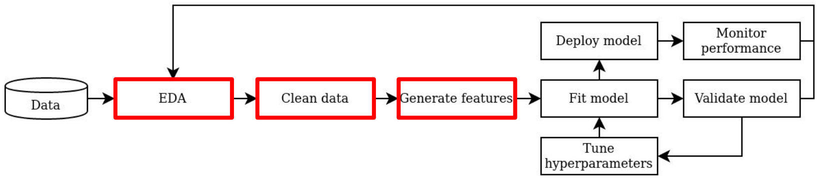
EDA
Cargo types

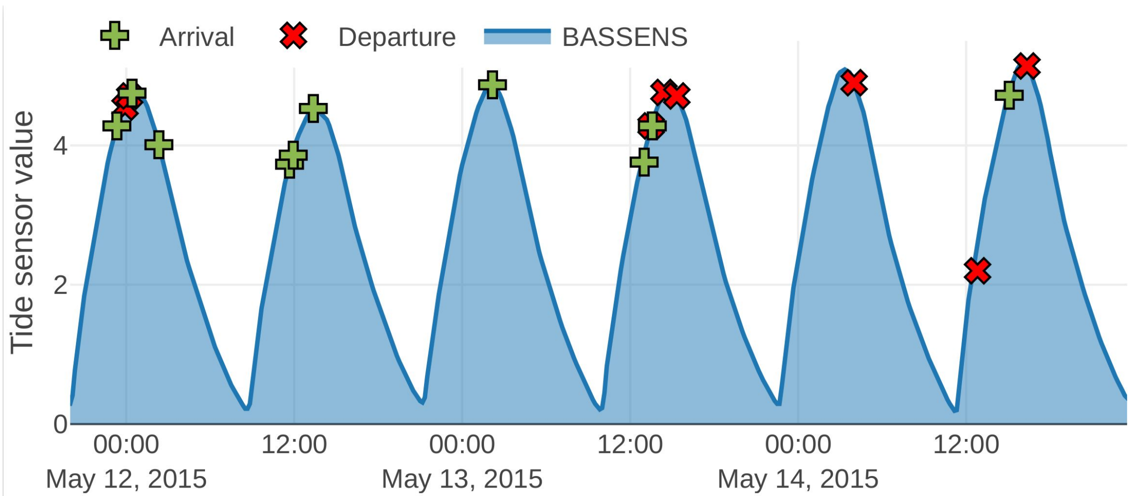


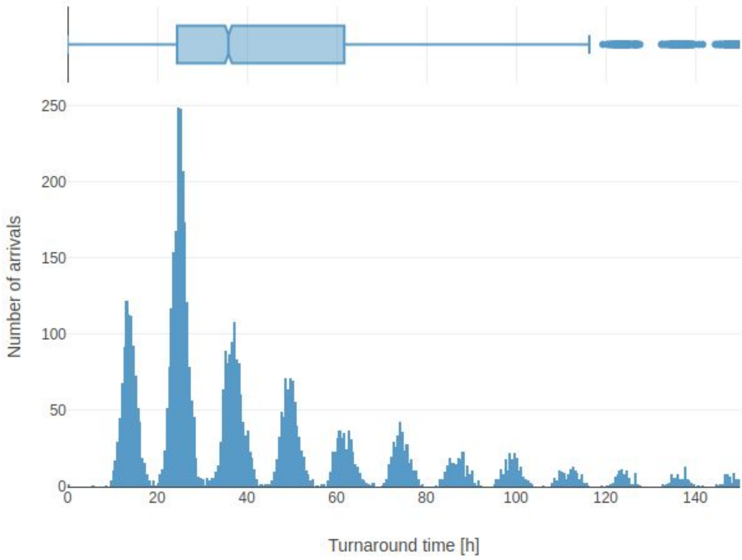


EDA

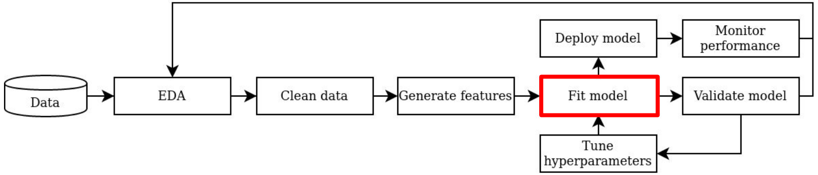
Water height - Tides







Model & experimental setup



- **Gradient boosting** (ensemble of weak tree based models)
- **CatBoost** python library (https://catboost.ai/ - Yandex)
  - CatBoost can handle categorical features
  - SOTA in comparison with other methods (XGBoost, RF, NN)
- **Cross-validation** 11 splits and final evaluation on 2 months live data
- MAE, RMSE, MAPE

Results (historical data)

Unloading				
	MAE [h]		RMSE	MAPE [%]
Unloading cargo type	CatBoost	Linear R.	CatBoost	CatBoost
BUTADIENE	**2.51**	4.46	4.33	13.87
METHANOL	**2.69**	6.18	4.25	8.51
SOYA OIL	**7.16**	8.47	13.28	22.49
CONTAINERS	**7.23**	8.42	9.79	22.81
RAPESEED OIL	**11.12**	11.68	17.84	22.81
SALT	**16.67**	19.3	25.4	41.15
BULK. FRETILIZERS	**17.49**	18.76	27.87	34.84
BULK UREA	**23.19**	26.7	32.03	37.07
NORTH SAWS	**24.87**	26.83	33.31	45.84
SUNFLOWER BULK	83.03	**82.37**	105.47	41.53
Top 10 cargo types (U)	**14.62**	16.33	27.46	28.43
All cargo types (U)	**15.75**	17.55	28.22	30.46

Loading				
	MAE [h]		RMSE	MAPE [%]
Loading cargo type	CatBoost	Linear R.	CatBoost	CatBoost
CONTAINERS	**7.24**	8.41	9.79	22.79
BULK CORN	**8.65**	10.51	13.89	29.64
PROD.CHIM.LIQ. AUTRES	**9.42**	11.48	15.58	32.78
BULK WHEAT	**10.04**	11.53	14.98	31.34
OTHER MINERALS	**12.12**	13.55	17.53	38.25
SUNFLOWER BULK	**13.34**	14.62	21.41	38.88
SUNFLOWER OIL	13.64	**13.54**	18.58	29.27
SCRAP	**22.6**	26.5	32.22	34.15
SUNFLOWER PELLETS	**23.06**	24.47	33.02	35.75
FAME	**23.61**	28.83	39.77	42.34
Top 10 cargo types (L)	**10.64**	12.37	17.93	29.53
All cargo types (L)	**11.57**	12.74	19.99	30.81

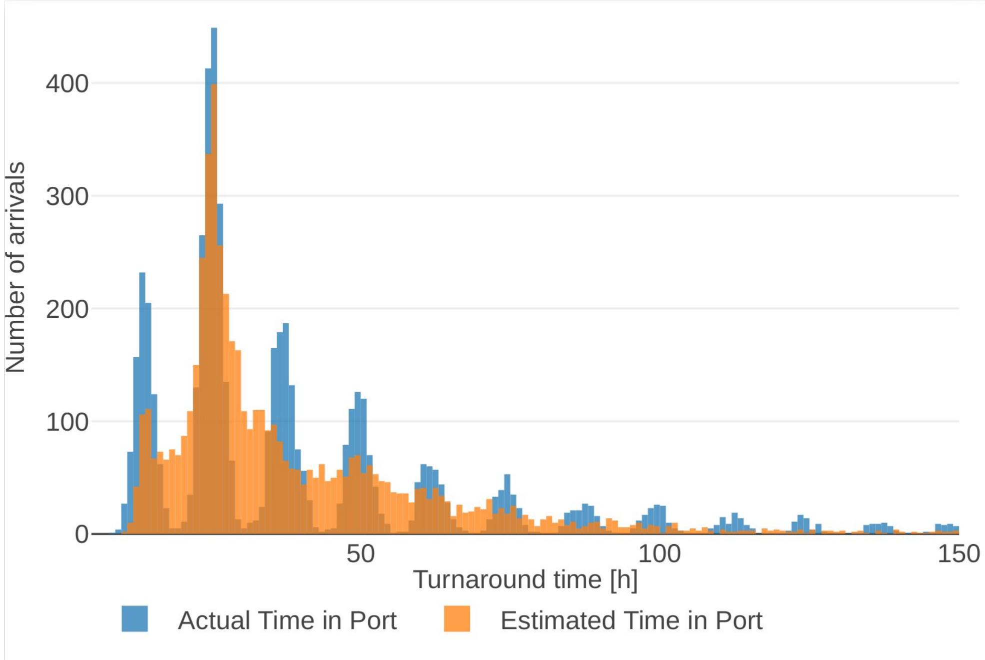
Results (live data)

Unloading		
Cargo type	CatBoost MAE [h]	Port MAE [h]
ENGR.LIQUIDES	**2.03**	18.66
METHANOL	**2.34**	9.49
RAPESEED OIL	**2.58**	32.16
BUTADIENE	**3.39**	16.73
SOYA OIL	**6.84**	24.74
TALL OIL	**7.32**	25.49
BULK UREA	**16.93**	57.29
NORTH SAWS	**25.37**	55.48
SUNFLOWER BULK	**89.77**	188.93
Combined	**16.52**	45.81

Loading		
Cargo type	CatBoost MAE [h]	Port MAE [h]
BULK CORN	**8.35**	19.41
CRUSHED TYRES	10.75	**9.75**
BULK WHEAT	**10.78**	17.72
SUNFLOWER OIL	**12.68**	14
SCRAP	**13.56**	56.88
Combined	**9.97**	22.75

Model analysis

Feature	Importance
cargo type (U)	17.4
cargo tonnage (U)	16.15
day of entry	12.71
berth (U)	11.13
cargo type (L)	8.54
hour of entry (round 4)	8.21
berth (L)	8.03
fiscal cargo type (L)	6.7
fiscal cargo type (U)	5.56
cargo tonnage (L)	4.72
holiday on entry	0.31
holiday in 2 days	0.2
holiday 1 day ago	0.18
holiday in 1 day	0.16



Conclusion

- We present first machine learning turnaround time prediction model based on FAL forms
- We evaluated it on 11 years of historical data and on live data
- It performs better than current manual system used in GPMB
- Integrating port specific equipment and HR information (future work)

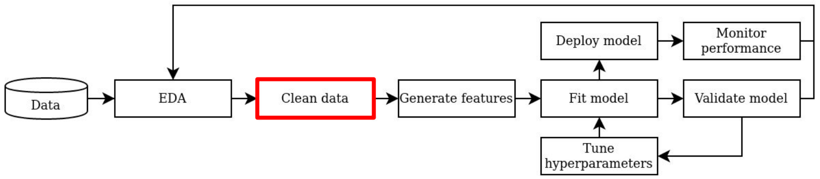
XLAB

Get IT done.

Port of Bordeaux (GPMB)

Bassens

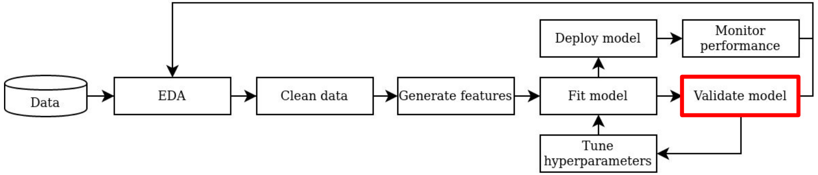
GPMB - ETD
Data cleaning

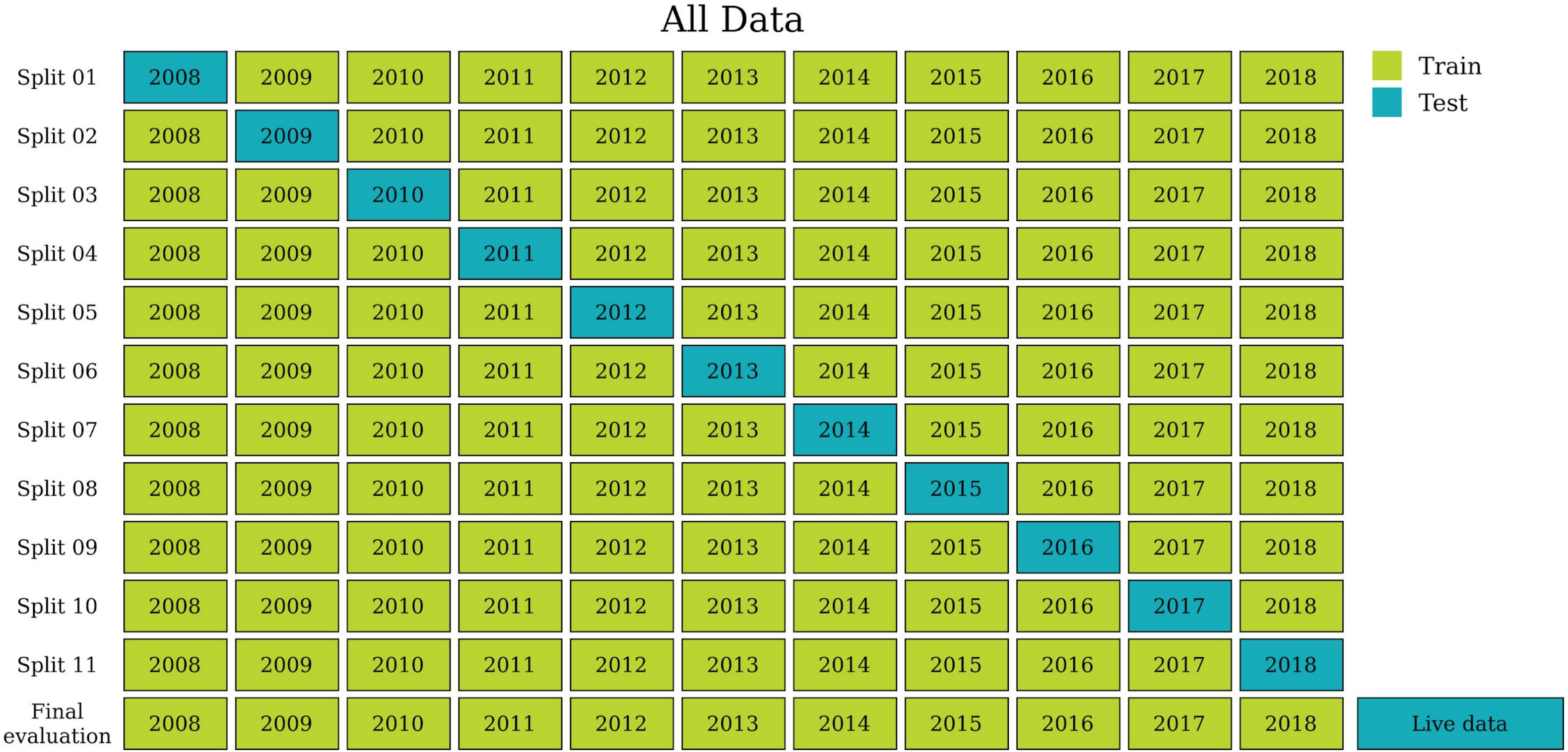


- Has cargo
- Not on maintenance berth
- “Normal” turnaround time
  - At least 1 h
  - Within 2 x std. from the median for cargo type
- At least 5 arrivals with same cargo types combination

GPMB - ETD

Cross-validation





Training and prediction times

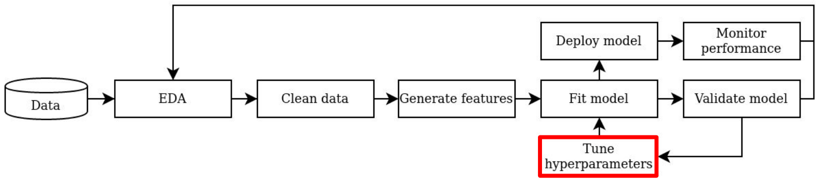
- Training of each fold takes about **35 seconds**
- Predictions are **almost instant**

Hardware:

- Intel i7-6700 CPU @ 3.40GHz
- Nvidia GeForce GTX 1050 Ti
- 32 GB RAM
- SSD

GPMB - ETD

Hyperparameters tuning



GridSearchCV method:

- Checks all combinations of passed hyperparameters (number of trees, learning speed, max trees depth ...)
- Using cross validation to find the best combination of hyperparameters