Dear Solvita,

I have attached 4 files (2 scripts, public\_data.csv and train\_input.csv). Below is the instruction to use our model.

@Yeldos @Rudolfs @Viesturs – do you have any comments on that? Maybe you could try and run it once?

@Solvita – if we do not have any comments from others, can you please submit it after 15.00? Please put me in cc: isurikovs@gmail.com

Description:

The model uses RandomForestRegressor algorithm to make predictions.

We have 2 script files: 1) **court\_cases.py**; 2) **data\_setup.py**.

1. **court\_cases.py** – main script where the model is trained, and predictions are done
2. **data\_setup.py** – module containing functions to process/transform data and extract some features (we made it a separate script just for convenience)

We have collected the data on number of judges in Latvian courts. Such statistics are publicly available on court websites. We saved all this data in **public\_data.csv** file. This file is used to get some additional features like court productivity.

Instructions:

1. Put in one folder: 1) **court\_cases.py**; 2) **data\_setup.py**; 3) **public\_data.csv**.
2. In the same folder put training dataset and rename it to **train\_input.csv** (just in case this file is attached to email).
3. Then put in this folder the file that is used for testing and name it as **input\_data\_to\_predict.csv**
4. Open the **court\_cases.py** file with IDE like Spyder or PyCharm and run it
5. After it finishes, it will create **predictions.csv** file. This file will contain all the data from the test file + created features (columns) by our model and the last column ‘predictions’ will have a prediction for each row.

Options (parameters in court\_cases.py):

* start\_date\_is\_used = True or False (line 55). By default it is True (it should be True for this competition). However, for new case predictions it will not be available. In that case change it to False.
* case\_end\_year\_in\_test\_data\_is\_only\_2018 = True or False (line 56). By default it is True (it should be True for this competition). If cases in the test set end not only in 2017, or if the model is used in production, then it should be changed to False.

Requirements:

* Our model is written in Python 3 (3.6)
* The following packages/libraries are used:
  + sklearn (scikit-learn)
  + pandas
  + numpy
  + datetime
  + dateutil

Regards,

Ilja