Predicting and Mitigating Emergency Situations on the roads: a Data-Driven Approach

Agenda

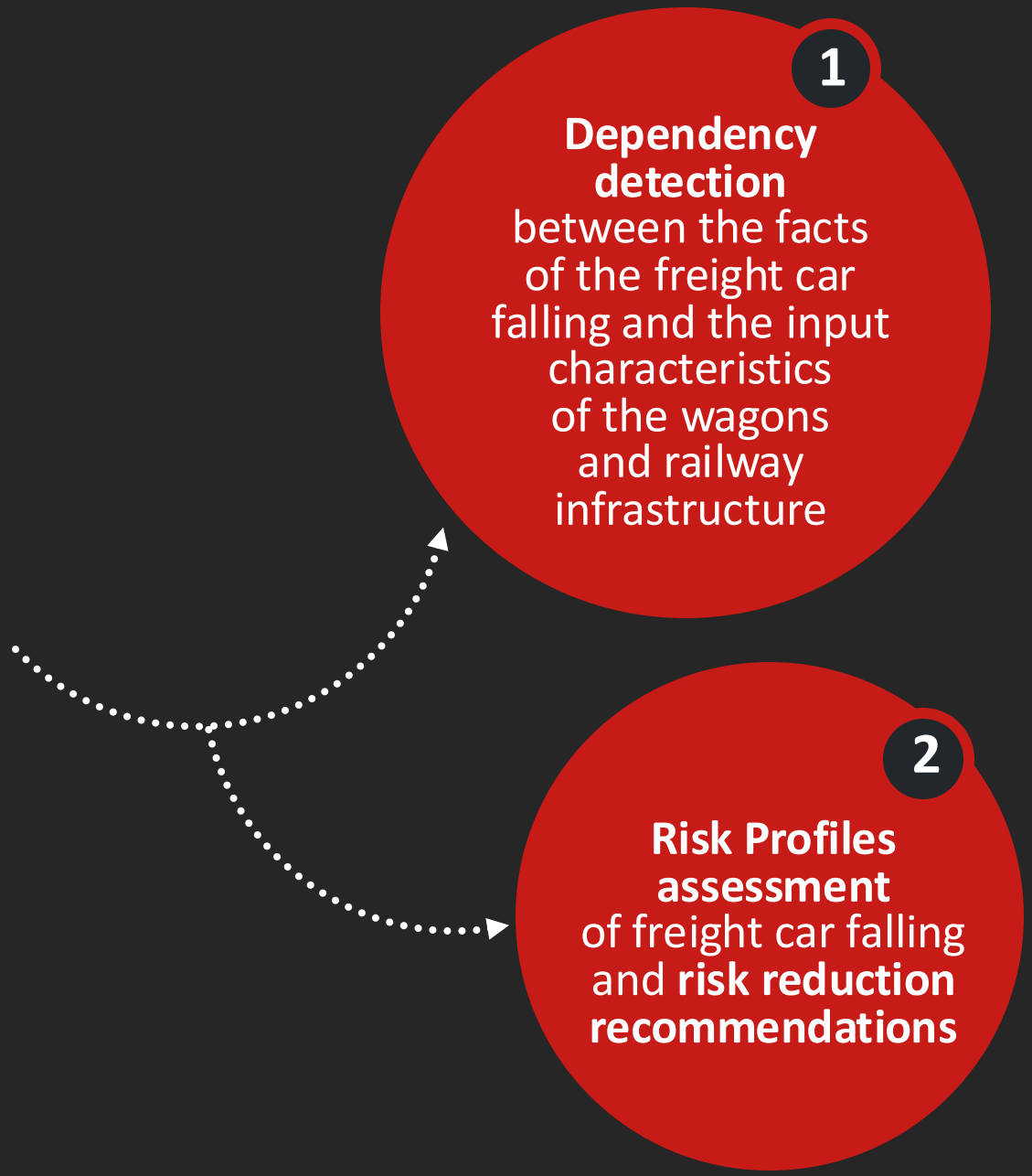
1. The problem is …
2. Action Plan and approach
3. It’s all about data
4. Model
5. Results

What is the problem?

- Frequent derailment incidents involving freight cars
- High repair costs
- Limited information on predictive factors
- Impact of accidents on business operations, safety, and the environment

Problem solving plan

- Preparing of historical data and a target variable for analysis
- Assessment of data quality for modeling
- Mathematical model for predicting the probability of freight car falling based on data (and risk profiles)
- Results of the project



Parameters:

Data for the period: Jan 2016 – July 2019

Data Category	Description	Source
Location Data	Wagon linked to station and time	Location from ASOUP, dispatches from TSO-31
Wagon Characteristics and Repair Data	Description of wagons and their repairs	ICH VP. Repairs
Derailment Data	List of derailments, enriched with infrastructure and wagon characteristic data	Derailment register from Excel by TCS
Location Data	Wagon linked to track	ASOUP-2V
Track Incident Data	List of track malfunction data, date of resolution, type of malfunction	ASUI
Repair Schedule Data	List of repair windows	APVO
Defect Drawing Data	Detailed description of defects	ASUI
Manifest Data	List of wagons in the train, wagon weight and cargo	ASOUP-2V
Wagon Passport Data	List of detailed technical characteristics	ABD PV
Speed Restriction Data	List of imposed restrictions and reasons for their implementation	ASUVOP 2
Weather Data	Weather conditions	System with data from weather stations
Station Coordinates Data	Geographical location of the station	GIS Railways

78

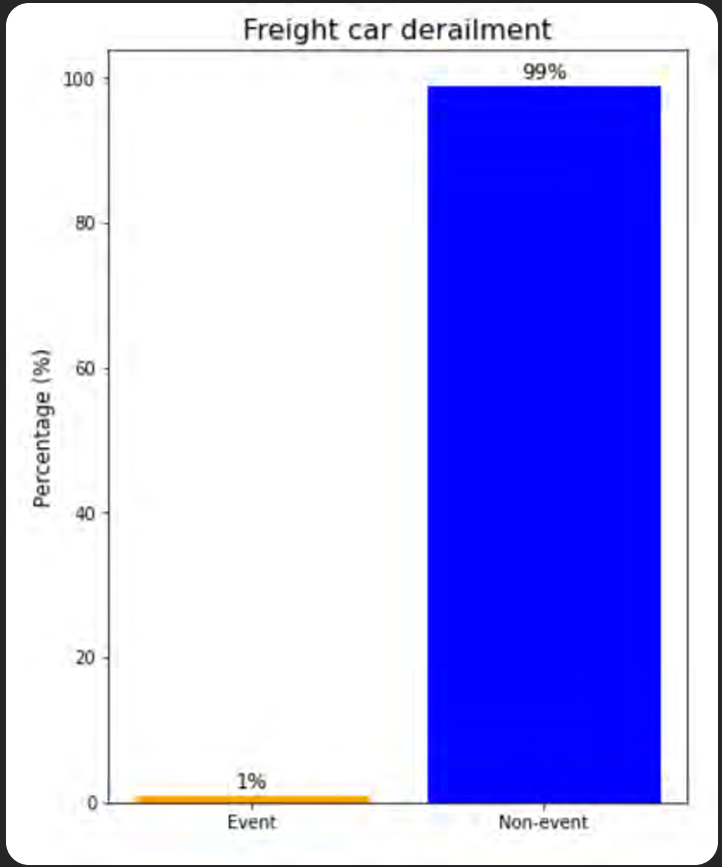
parameters

Model

The challenge №1: → 2 approaches → Reduce the number of non-events (we are losing information)

Bad practice

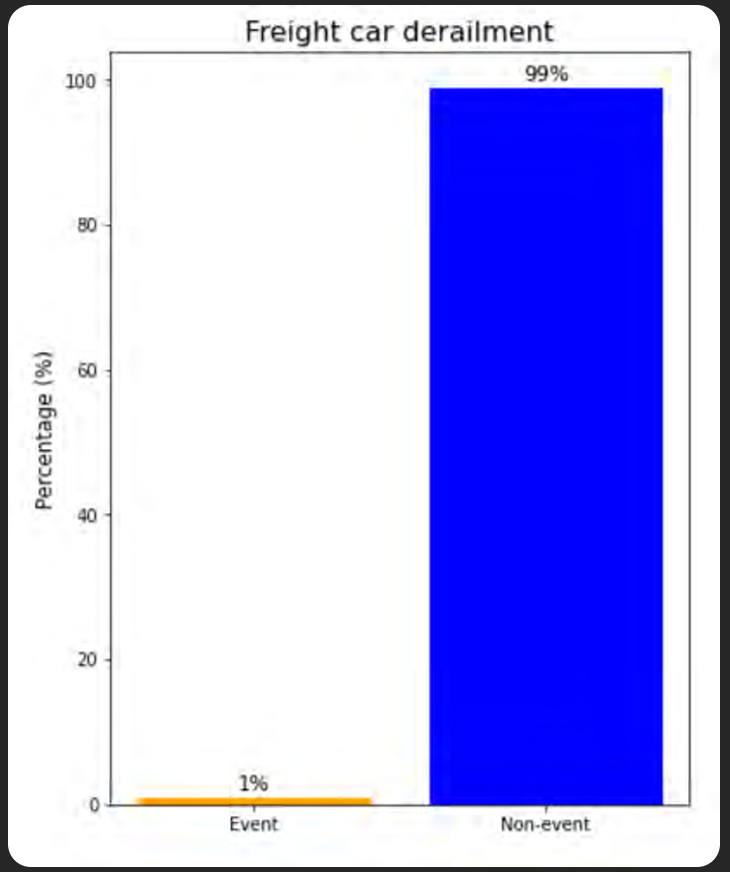
1% event to 99% non-event



Model

The challenge №1:

1% event to 99% non-event



2 approaches → Reduce the number of non-events

→ Dealing with class imbalance

Bad practice

Best idea

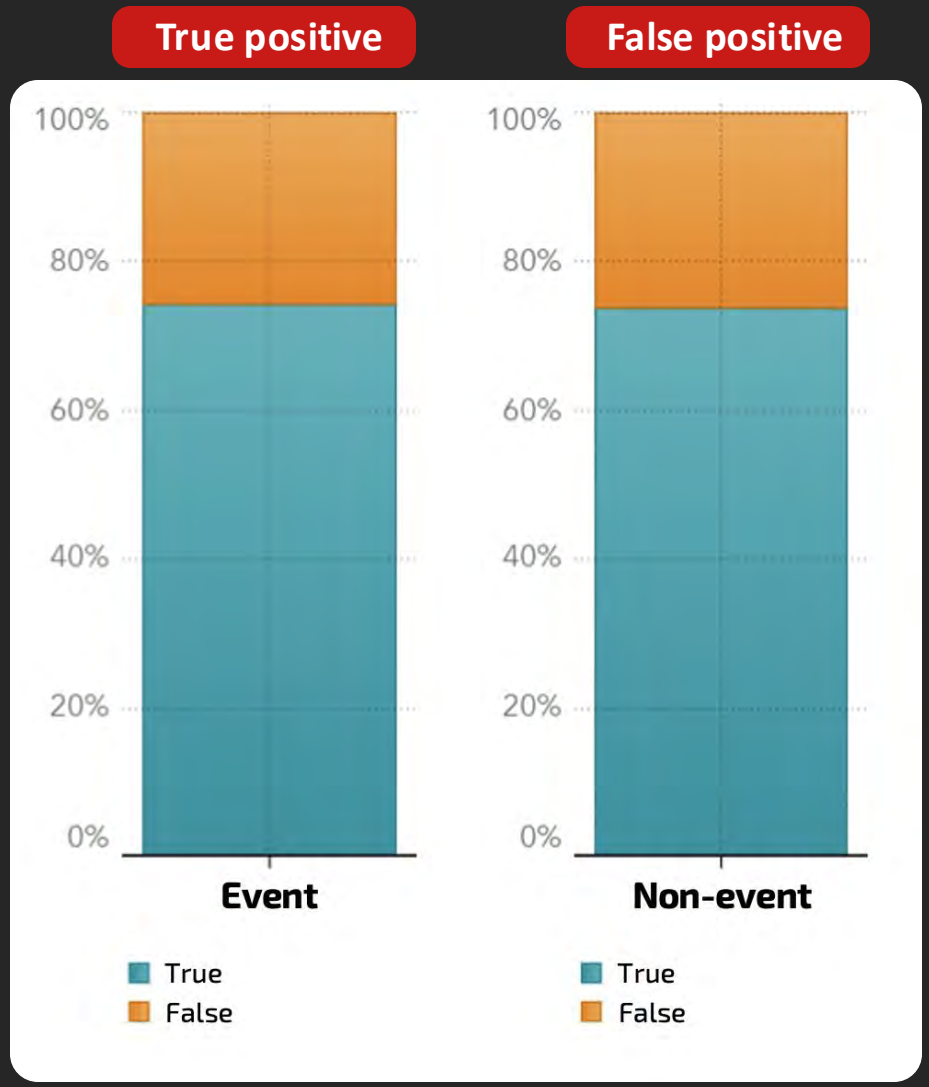
SMOTE

will allow us to better account for our events in the prediction model

Model

Algorithm:

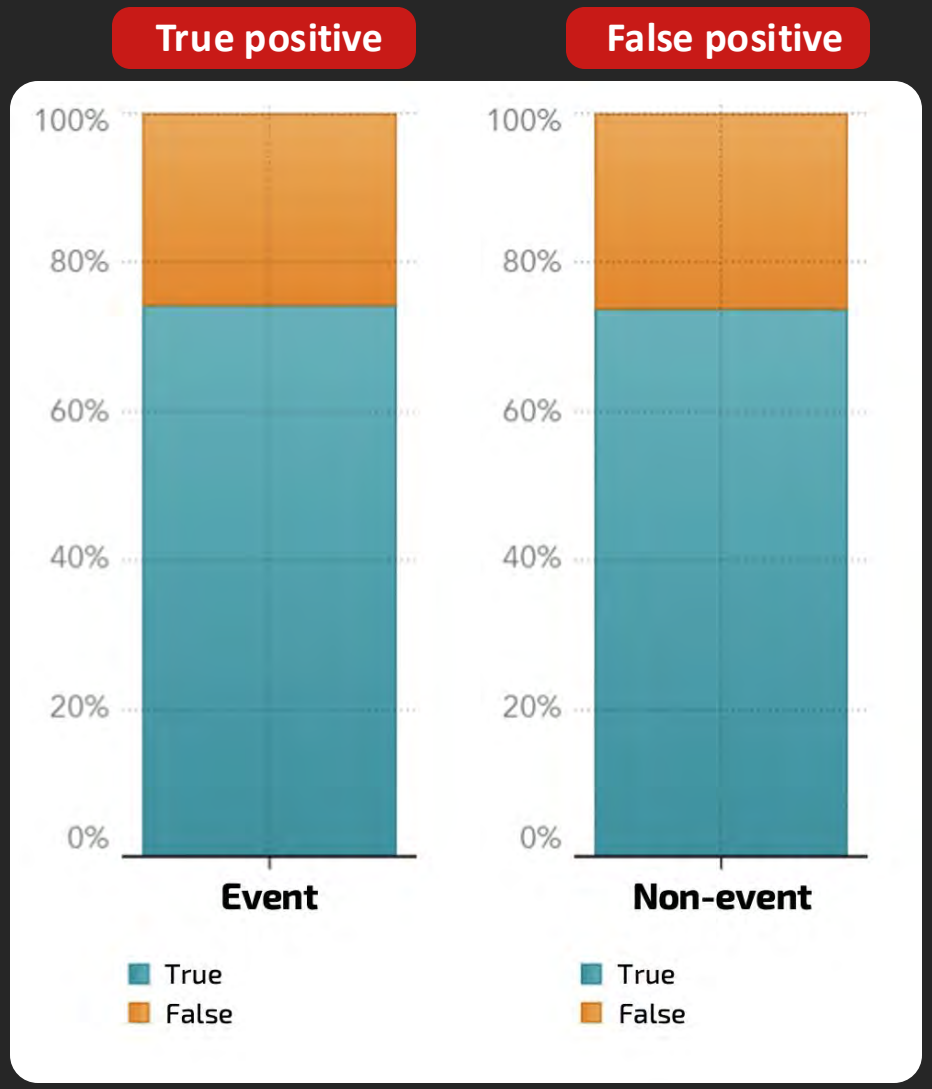
Random Forest →

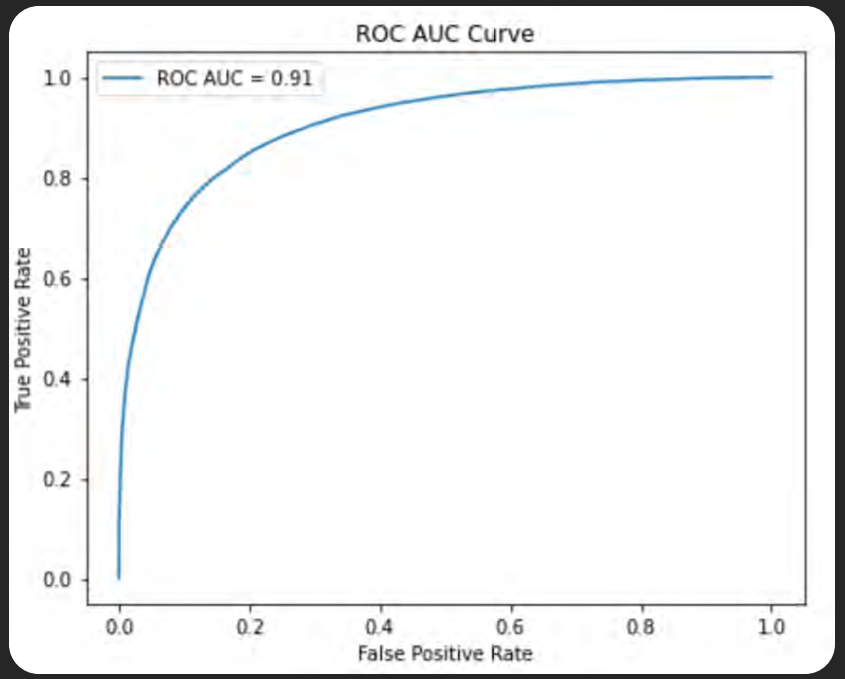


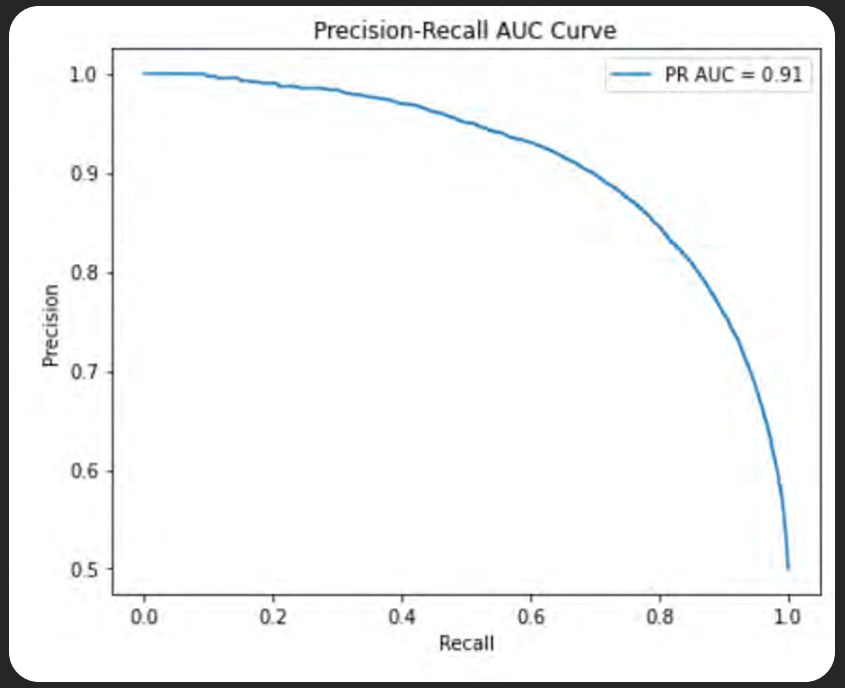
Model

Algorithm:

Random Forest

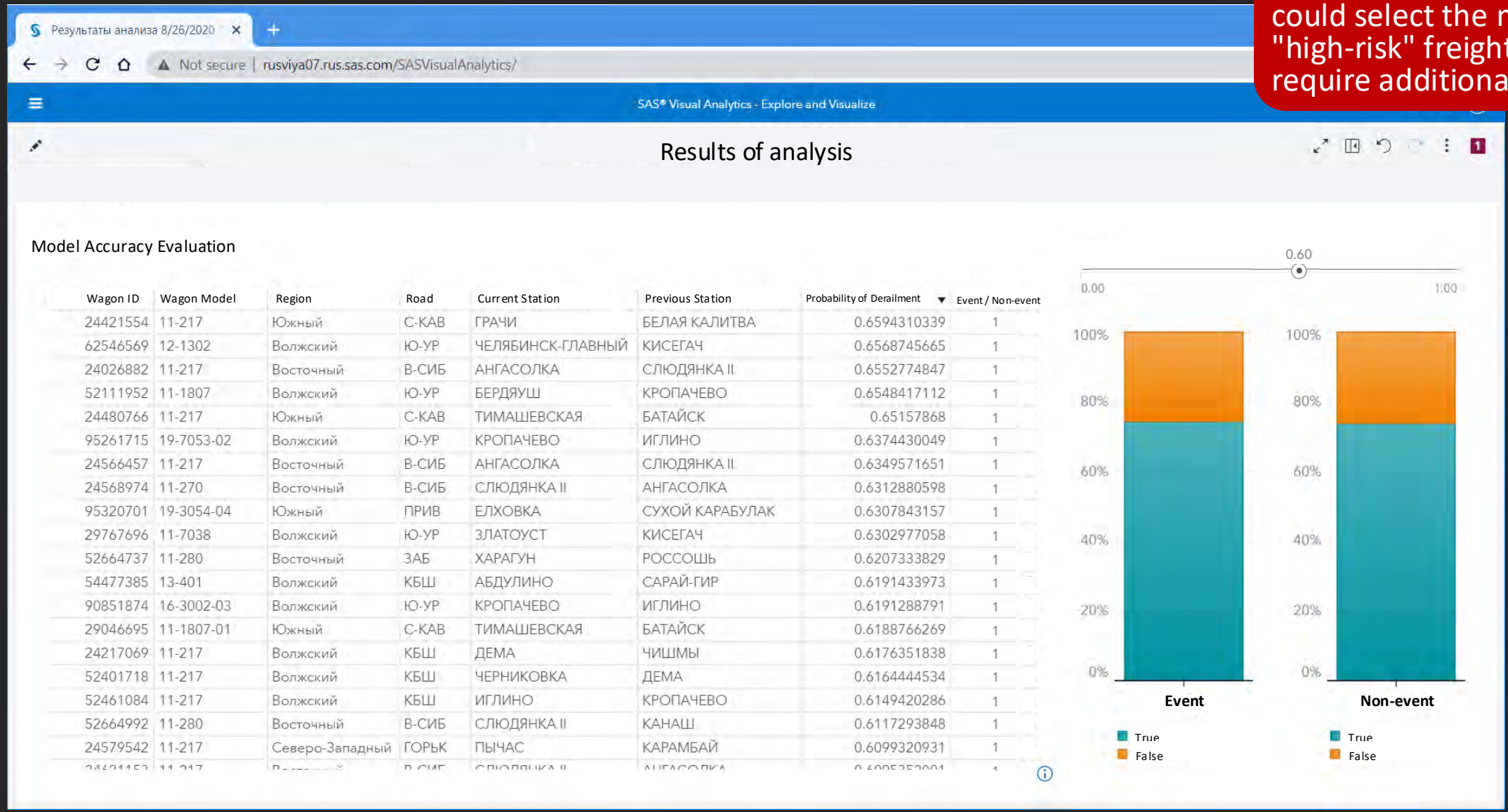






Probability of freight car derailment

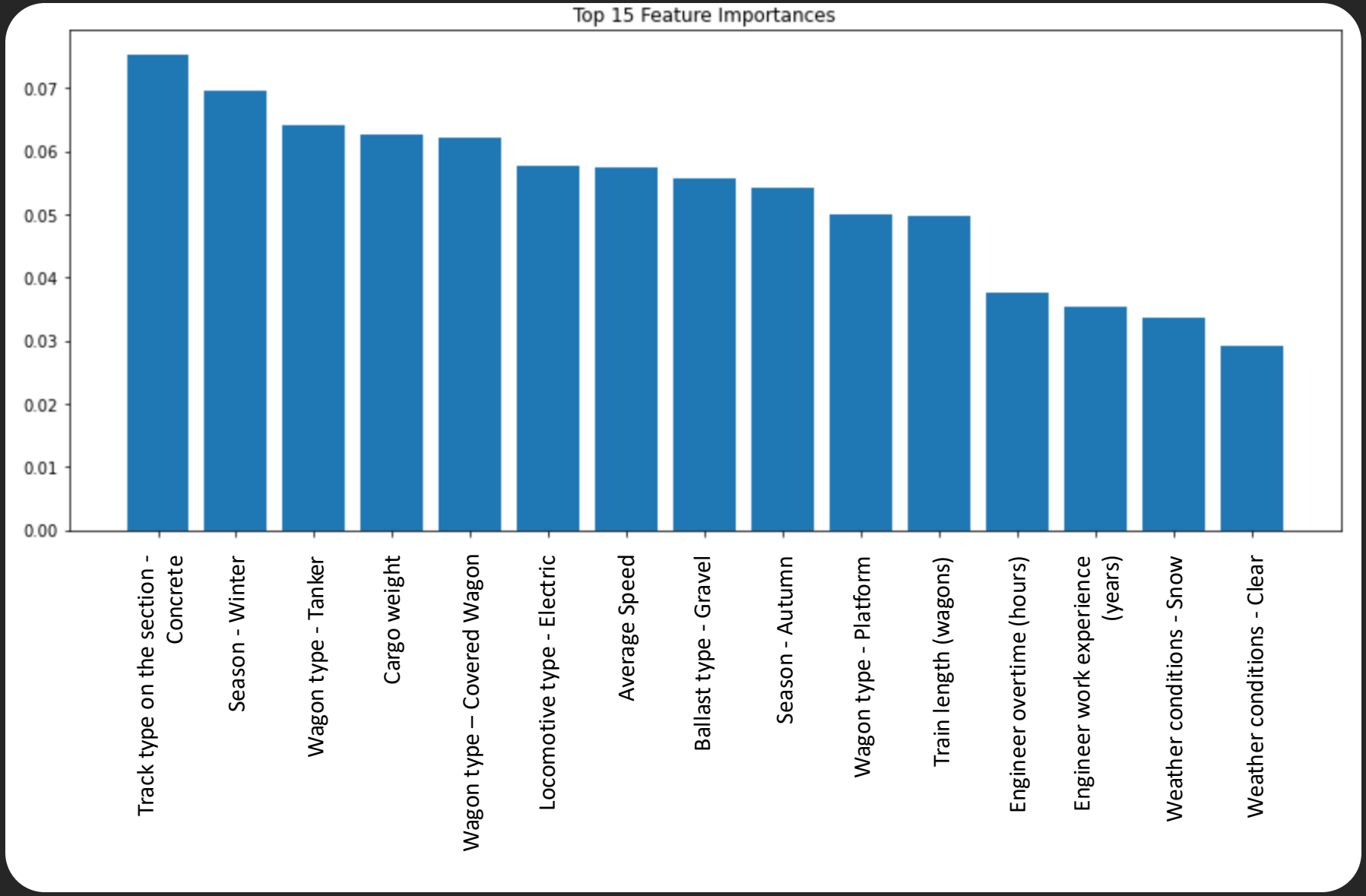
Based on the probability, we could select the most "high-risk" freight cars that require additional inspection



Wagon ID	Wagon Model	Region	Road	Current Station	Previous Station	Probability of Derailment	Event / Non-event
24421554	11-217	Южный	С-КАВ	ГРАЧИ	БЕЛАЯ КАЛИТВА	0.6594310339	1
62546569	12-1302	Волжский	Ю-УР	ЧЕЛЯБИНСК-ГЛАВНЫЙ	КИСЕГАЧ	0.6568745665	1
24026882	11-217	Восточный	В-СИБ	АНГАСОЛКА	СЛЮДЯНКА II	0.6552774847	1
52111952	11-1807	Волжский	Ю-УР	БЕРДЯУШ	КРОПАЧЕВО	0.6548417112	1
24480766	11-217	Южный	С-КАВ	ТИМАШЕВСКАЯ	БАТАЙСК	0.65157868	1
95261715	19-7053-02	Волжский	Ю-УР	КРОПАЧЕВО	ИГЛИНО	0.6374430049	1
24566457	11-217	Восточный	В-СИБ	АНГАСОЛКА	СЛЮДЯНКА II	0.6349571651	1
24568974	11-270	Восточный	В-СИБ	СЛЮДЯНКА II	АНГАСОЛКА	0.6312880598	1
95320701	19-3054-04	Южный	ПРИВ	ЕЛХОВКА	СУХОЙ КАРАБУЛАК	0.6307843157	1
29767696	11-7038	Волжский	Ю-УР	ЗЛАТОУСТ	КИСЕГАЧ	0.6302977058	1
52664737	11-280	Восточный	ЗАБ	ХАРАГУН	РОССОШЬ	0.6207333829	1
54477385	13-401	Волжский	КБШ	АБДУЛИНО	САРАЙ-ГИР	0.6191433973	1
90851874	16-3002-03	Волжский	Ю-УР	КРОПАЧЕВО	ИГЛИНО	0.6191288791	1
29046695	11-1807-01	Южный	С-КАВ	ТИМАШЕВСКАЯ	БАТАЙСК	0.6188766269	1
24217069	11-217	Волжский	КБШ	ДЕМА	ЧИШМЫ	0.6176351838	1
52401718	11-217	Волжский	КБШ	ЧЕРНИКОВКА	ДЕМА	0.6164444534	1
52461084	11-217	Волжский	КБШ	ИГЛИНО	КРОПАЧЕВО	0.6149420286	1
52664992	11-280	Восточный	В-СИБ	СЛЮДЯНКА II	КАНАШ	0.6117293848	1
24579542	11-217	Северо-Западный	ГОРЬК	ПЫЧАС	КАРАМБАЙ	0.6099320931	1

Variable Importance

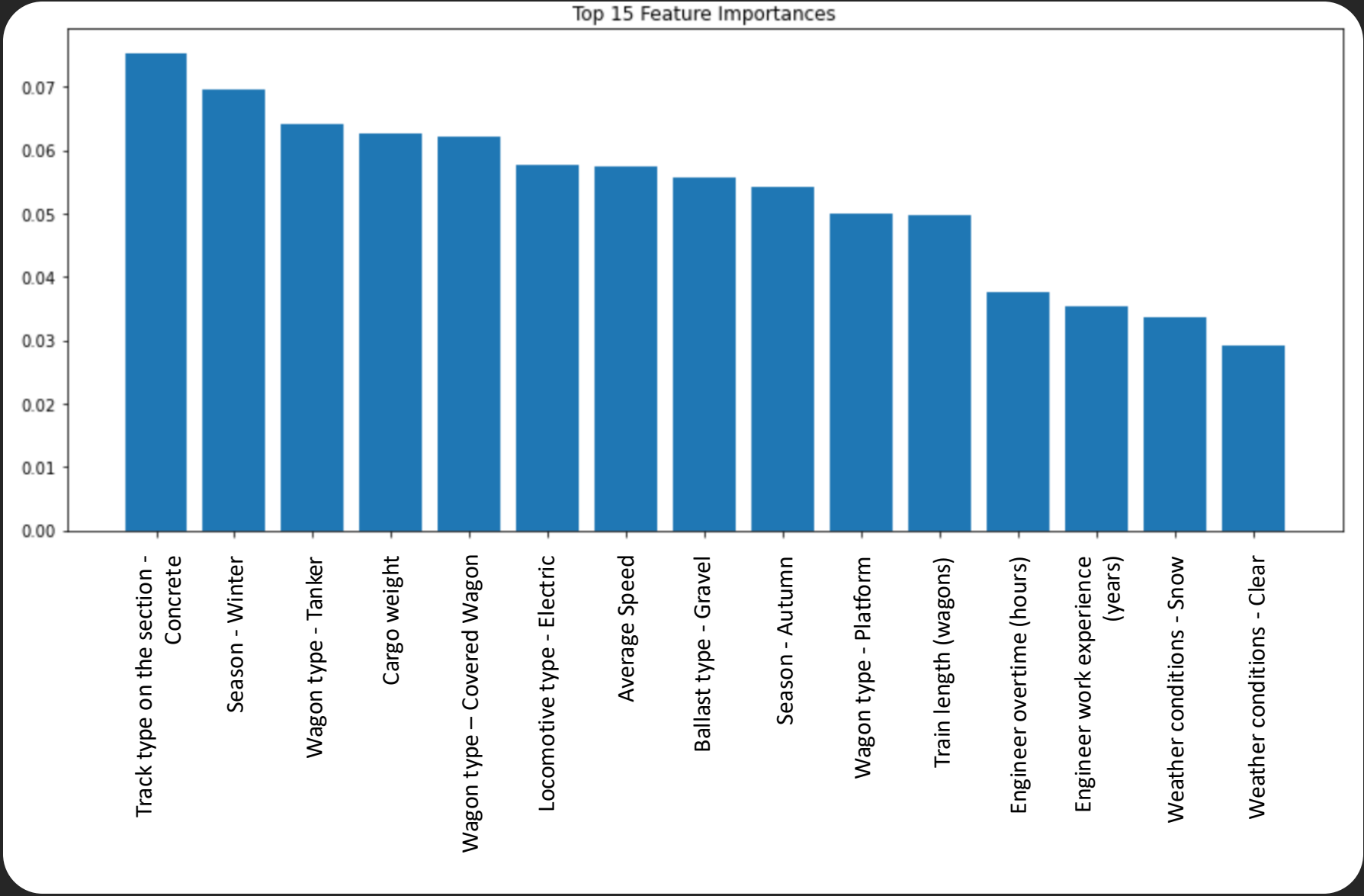
But these are still not risk profiles



Variable Importance

The challenge №2

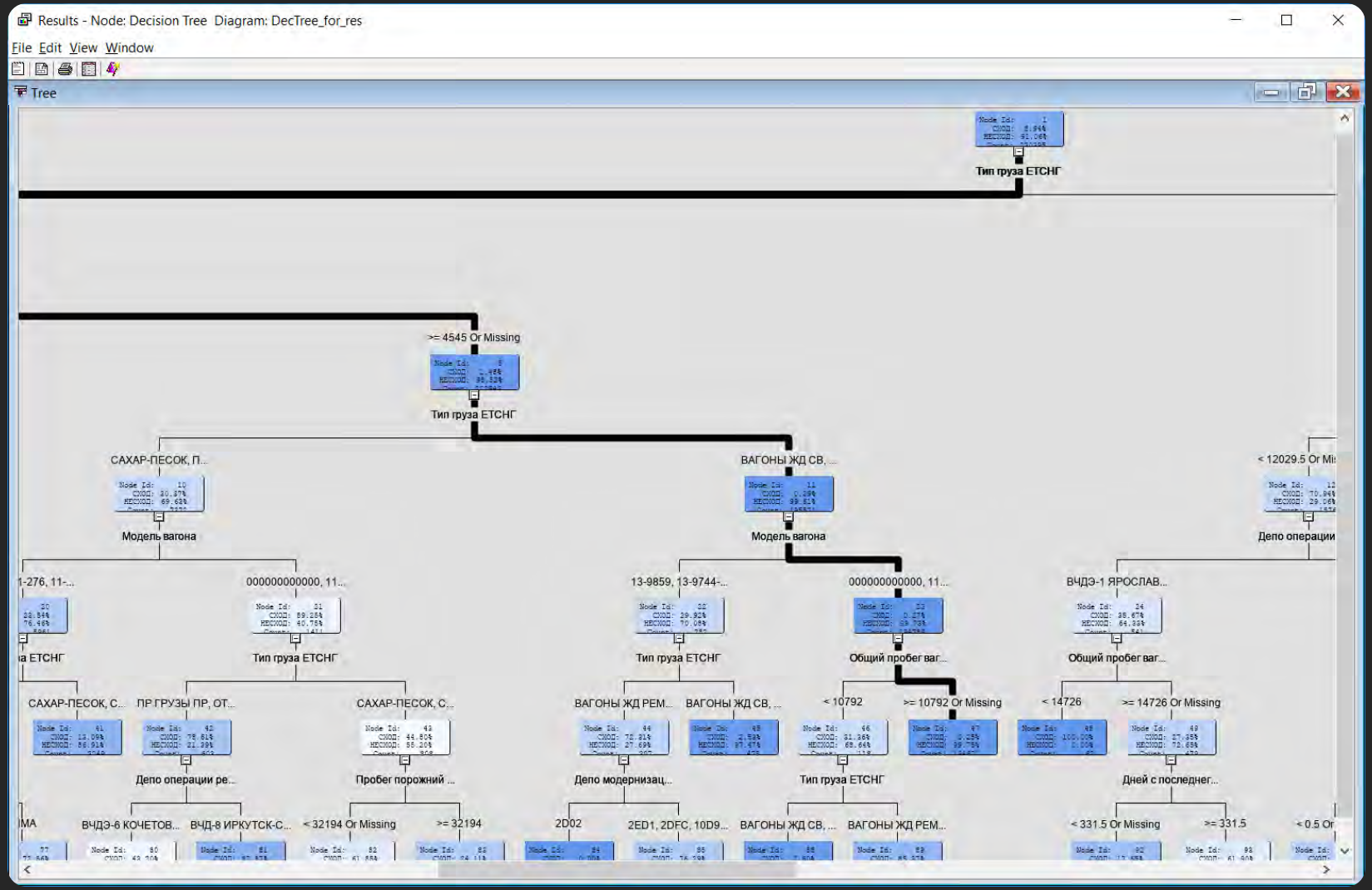
But these are still not risk profiles



Risk Profiles

Let’s build a Decision Tree based on Random Forest results

OMG



Risk Profiles

- IF Number of wagons < 38.5 AND Years from issue < 51.5 AND Ratio: speed on the last section to average speed < 1.0385
  94,5% Probability (Share 21%)

- IF Number of wagons >= 38.5 AND Type of dispatch «Group of Wagon» AND Condition (Wagon Weight) = Empty AND Train Weight >= 175425 AND Ratio: speed on the last hauls to average speed >= 0.94
  78,6% Probability (Share 15%)

- IF Остаток ресурса в км до нормы >= 1014.5 AND Кол-во вагонов >= 38.5 AND Type of dispatch «Group of Wagon» AND Condition (Wagon Weight) = With Cargo AND Days to scheduled repairs >= 437.5
  99,9% Probability of **not fall** (Share 47%)

Project Results

80 %
reducing accidents

Minimizing the risks of environmental disasters

12 $ M
on repair costs savings per year