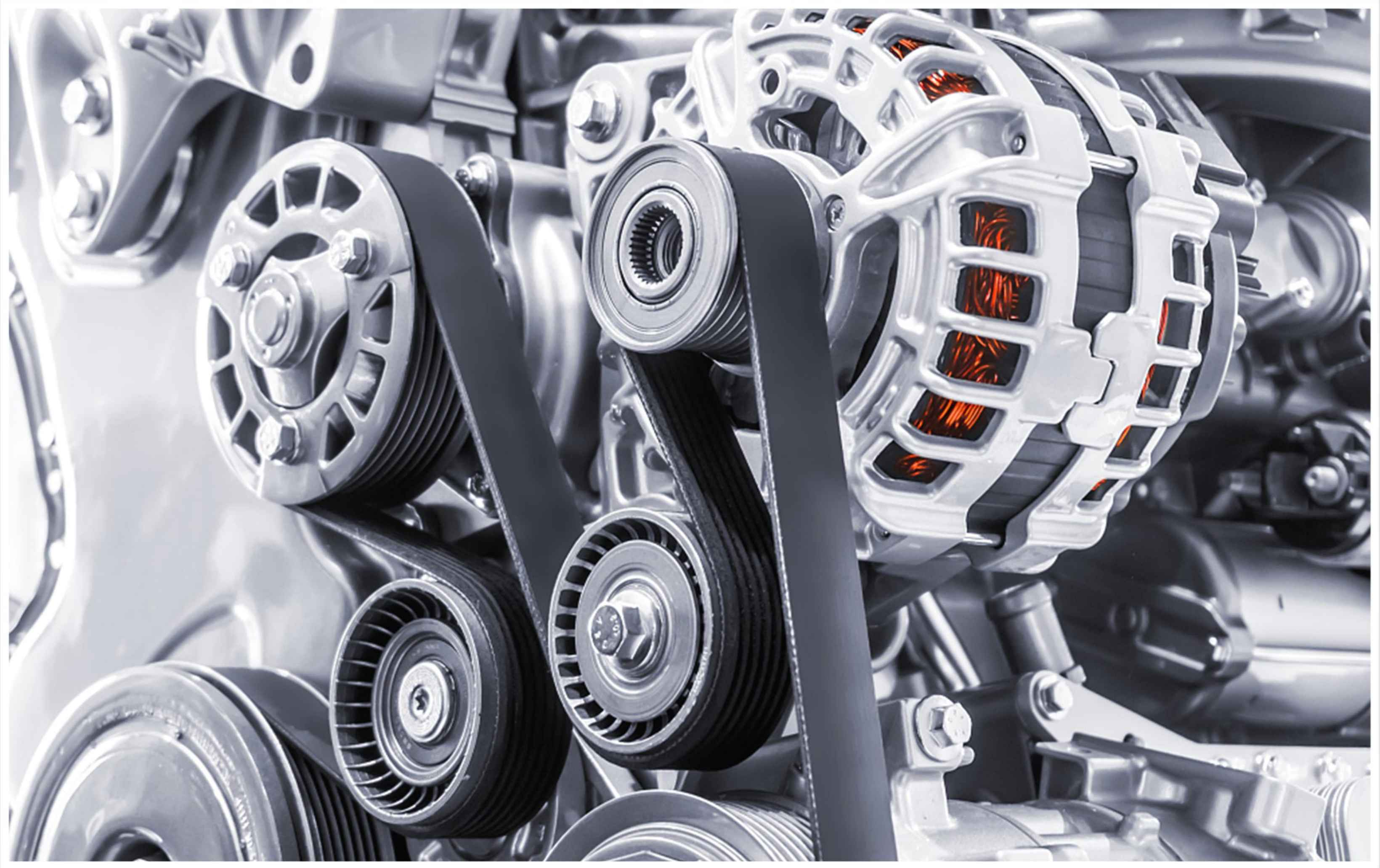
Engine Failure Prewarning Algorithm

Group 20-Zhaoyuan Qiu, Wenxuan Tang, Yiping Lu

2022/02/01-2022/4/20



Key Words: Engine States Space, Classification Model, Prewarning Threshold

I Data Overview

Data collected on the real-time operation of 4 freight vehicles belonging to Valin Xingma Automobile (Group) Co., LTD during the 12-month period in 2020 by Robert Bosch GmbH.

TM	DPFCarbonLoad	ECUMile	EngineSpeed	PTOSwitch	SCRUpstreamTemp	SCRDownstreamTemperature	T15Status	T4
2019-12-18 23:58:58.000	0.0	153821.250	1331.5	2.0	197.43750	218.93750	1.0	184.75000
2019-12-18 23:58:59.000	0.0	84657.875	649.5	0.0	151.46875	181.43750	1.0	129.06250

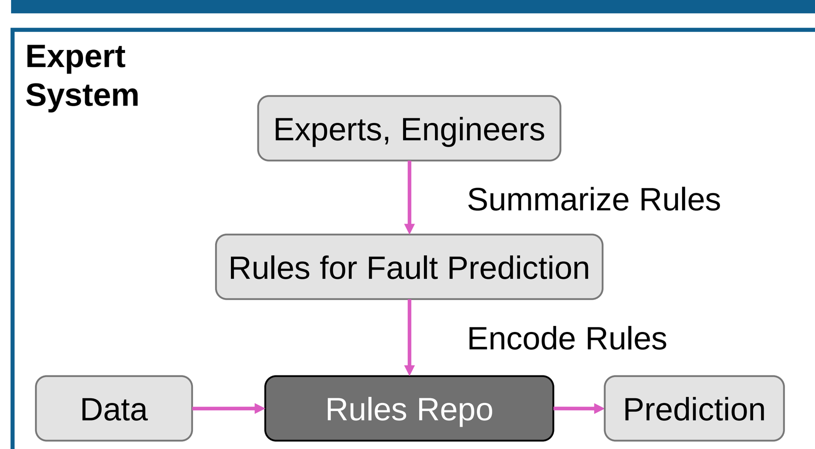
ActiveRegenerativeState	InternalTorque	RegenerativeRequest	Brake	ReferentialTorque	EngineOperatingTime	FuelInjectionQuantity	AtmosphericPressure
1048577.0	13.900000	2.0	0.0	1833.0	28589924.0	1.680000	1017.0
1048577.0	195.000000	3.0	0.0	1833.0	37094352.0	17.879999	1011.0

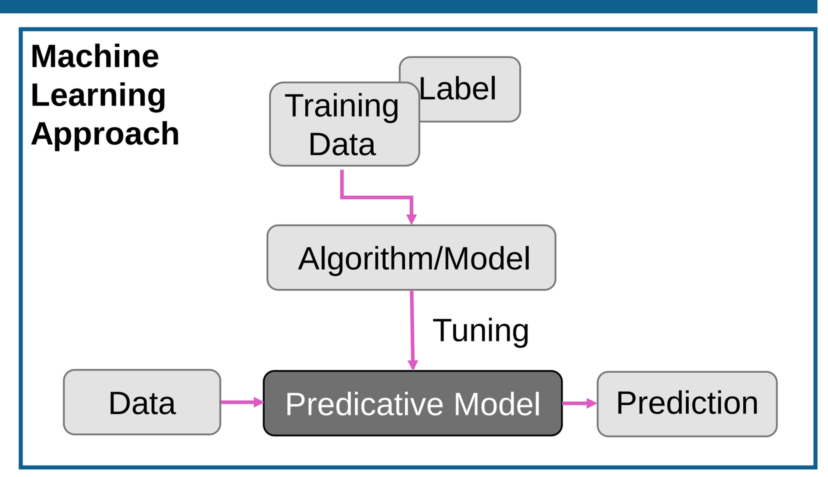
TotalCH_InjectionQuantity	AbsoluteBoostPressure	MeteringValveFlowControl	Speed	AirIntake	AirTemp	LimitedTorsionalActivationMode	SPN|FMI
0.0	114.0	840.0	58.347656	514.00	18.799999	0.0	Normal|Normal
0.0	102.0	1130.0	0.000000	228.00	38.599998	0.0	Normal|Normal

Data Glance

Valid variables: 50
Label: 'SPN|FMI'
Time granularity: second-level

II Classic Expert System VS Machine Learning





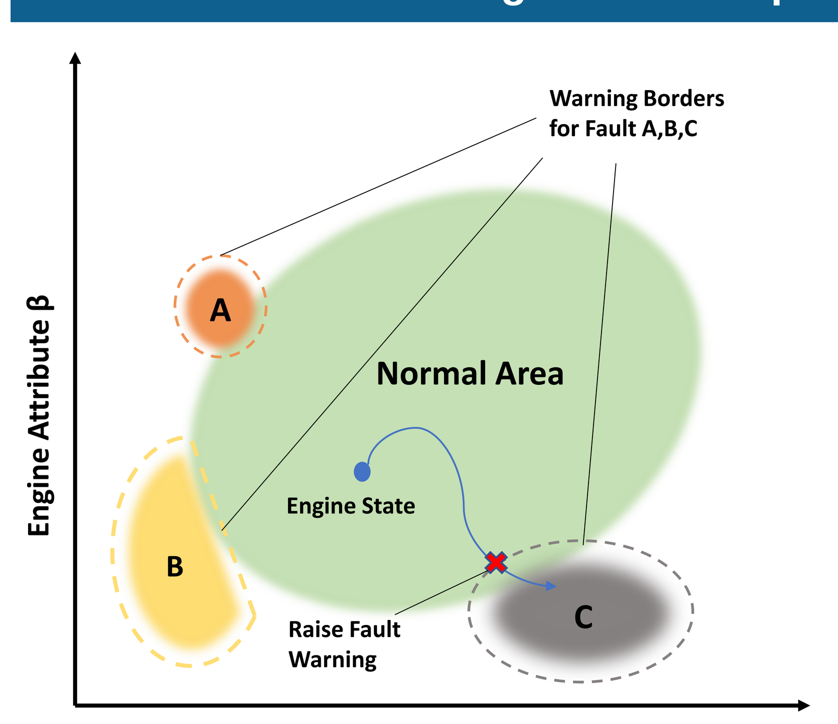
Expert System — Advantages
- System is highly stable.
- Rules are highly interpretable.

Expert System — Drawbacks
- Development demands expertise & experience.
- Hard-coded rules are hard to update.

Machine Learning — Advantages
- More independent on expertise & experience.
- Easy to update and make generalization.

Machine Learning — Drawbacks
- Greater demand for data quantity & quality.
- Model sometimes lose interpretability.

III Abstract Model of Engine States Space



- The set of all possible states construct an N-dimensional space determined by n attributes of the engine.
- Normal state and each fault state owns a distribution field.
- When engine operates, its state is a point of continuous motion in this space.
- When the engine state is in the fault field, a fault occurs.

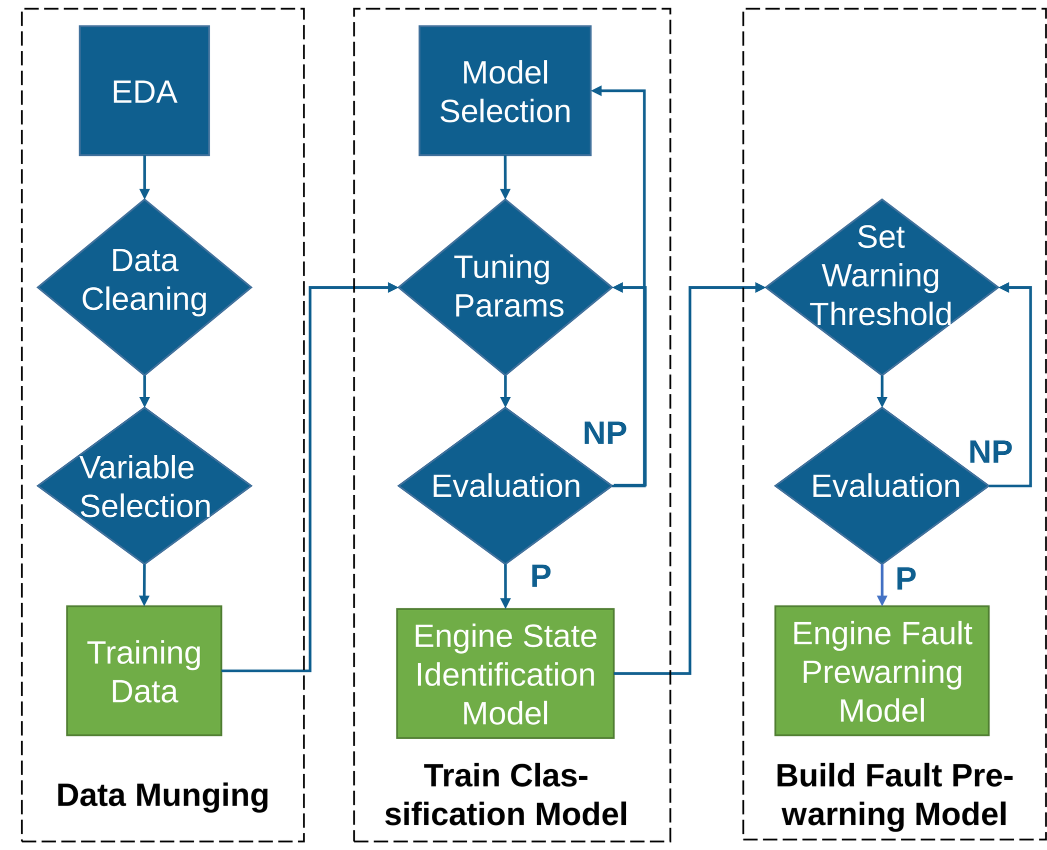
How to locate engine state in the space?

Classification model, which uses probability to measure the distance from each area.

When to make warnings?

The time when engine state approaches to fault area borders. Warning borders are to be built.

IV Project Work Flow



Part1 Data Munging

- Drop missing value and duplicates.
- Deal with time discontinuity.
- Select variable with RandomForest.

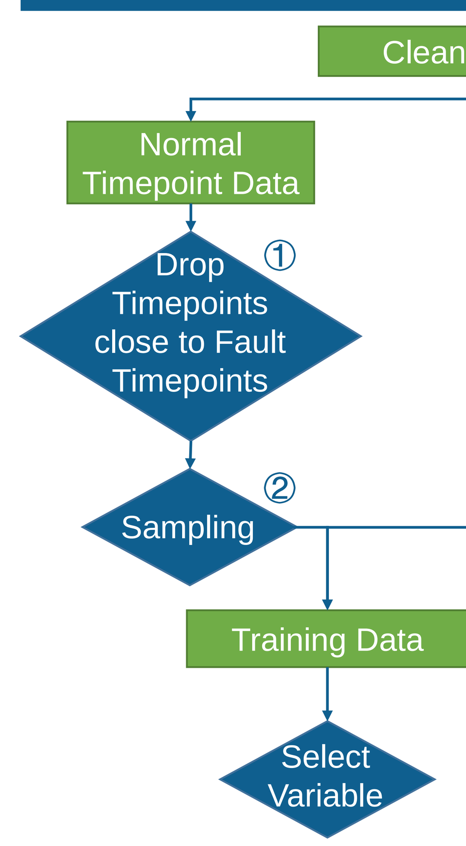
Part2 Train Classification Model

- RandomForest, KNN, XGBoost
- Cross Validation
- Ensemble method

Part3 Build Fault Prewarning Model

Find optimal prewarning threshold for each fault type.

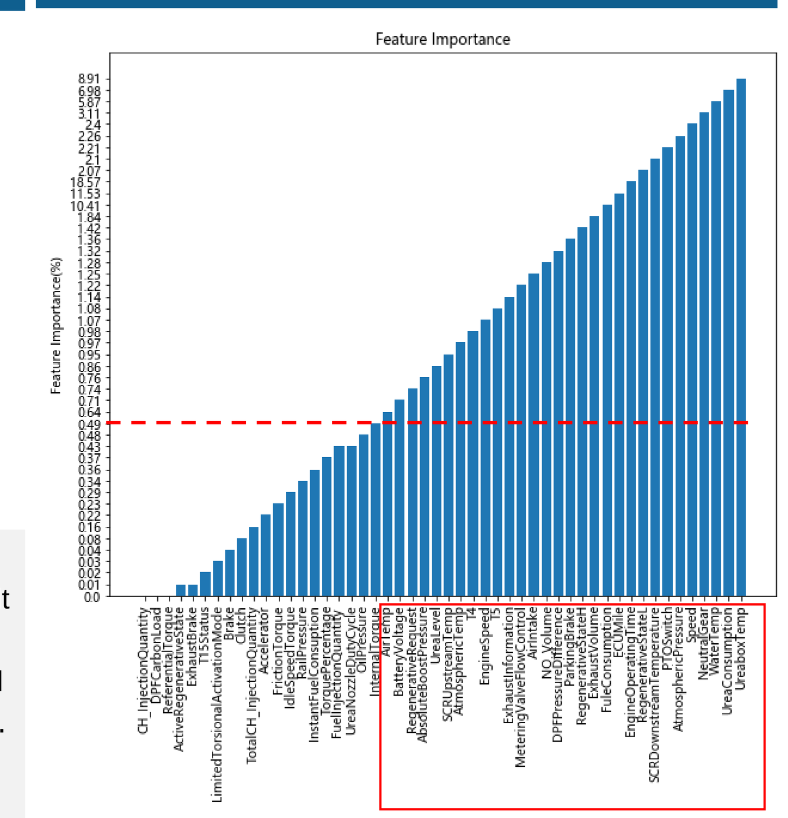
V Data Preparation for Variable Selection



Step Goals & Reasons

① Make normal data and fault data more distinguishable.
② Limit the number of normal data while drop close records.
③ Samples are too few to make model convincing.

Select Variable with Random Forest



VI Engine State Identification Model Training

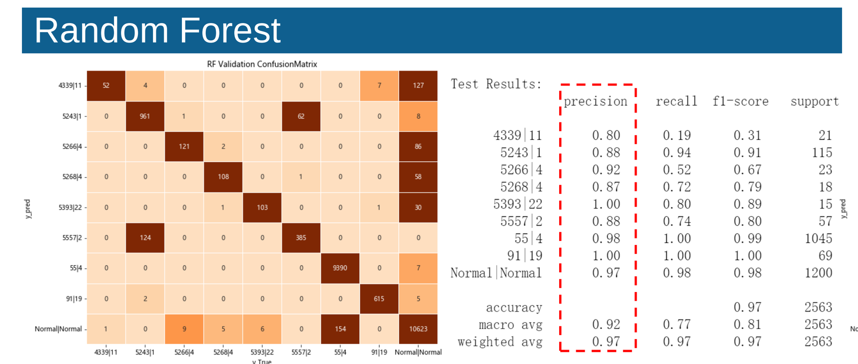
Evaluation Metrics

- Confusion Matrix
- Precision $P(Label = A|Prediction = A)$
- Recall $P(Prediction = A|Label = A)$
- F1 Score $2 \times (Precision^{-1} + Recall^{-1})^{-1}$
- Accuracy $P(Label = Prediction)$

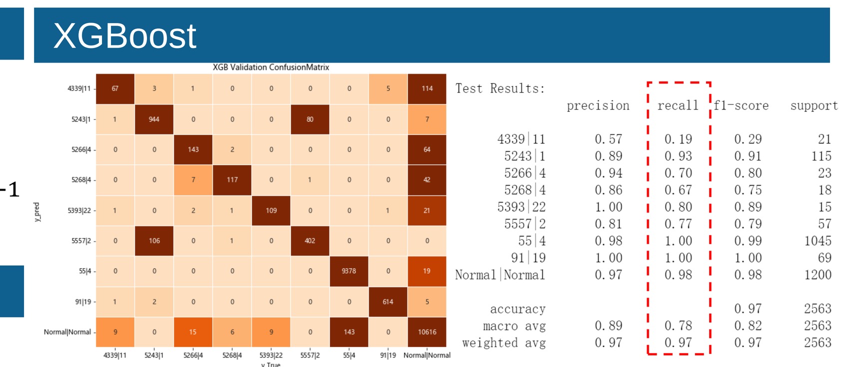
Methods for Tuning Hyperparameters

- Train-Test Split by 9:1
- 10-Folds Cross Validation
- Grid Search within Small Range

Random Forest

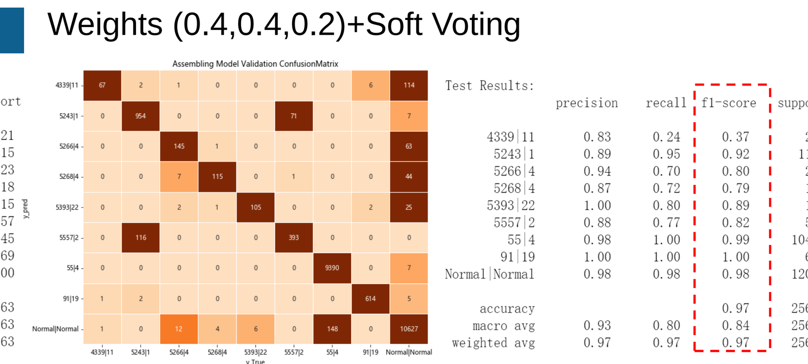


XGBoost



Random Forest-XGBoost-KNN Ensemble Model

Weights (0.4,0.4,0.2)+Soft Voting



References

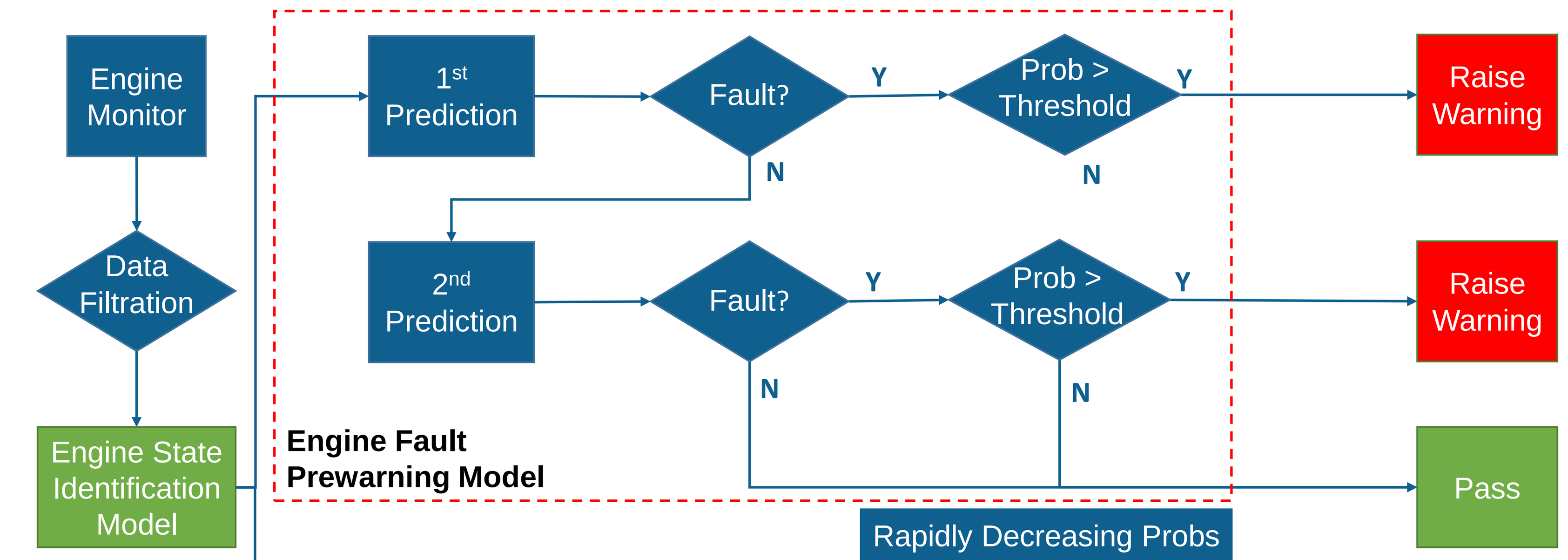
[1] Strobl C , Boulesteix A L , Zeileis A , et al. Bias in random forest variable importance measures: Illustrations, sources and a solution[J]. Bmc Bioinformatics, 2007, 8

[2] Friedman J H. Greedy Function Approximation: A Gradient Boosting Machine[J]. Annals of Statistics, 2001, 29(5):1189-1232.

[3] Chen T , Guestrin C . XGBoost: A Scalable Tree Boosting System[J]. the 22nd ACM SIGKDD International Conference, 2016.

[4] Aurélien Geron. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow [A]. O'Reilly Media, Inc., 2019：162-192

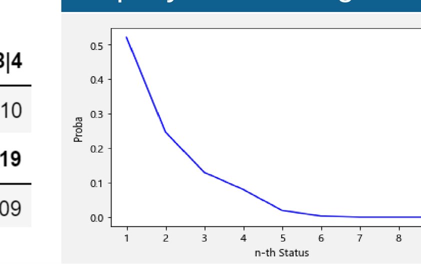
VII Fault Pre-warning Model



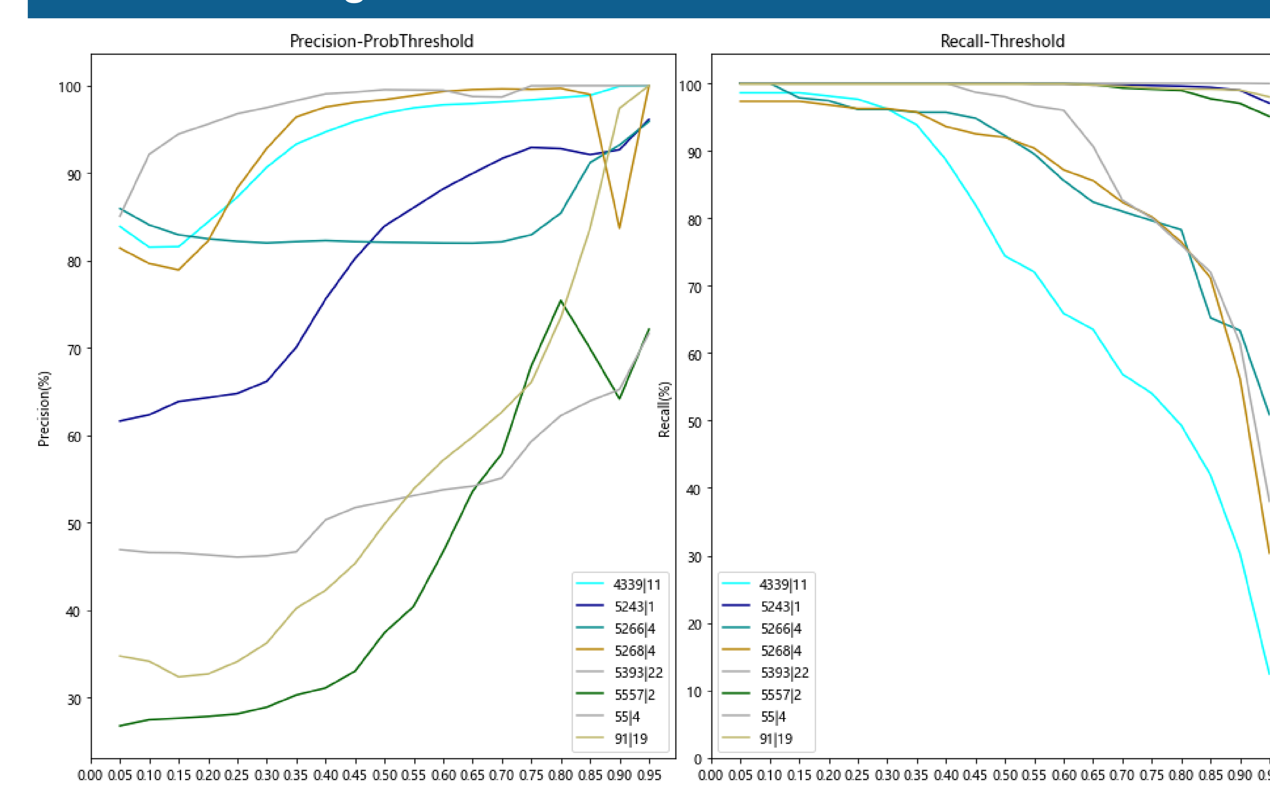
Probs of State

	4339|11	5243|1	5266|4	5268|4
	0.017782	0.248251	0.067288	0.003210
Normal|Normal	**5393|22**	**5557|2**	**55|4**	**91|19**
0.139540	0.000019	0.523882	0.000018	0.000009

Rapidly Decreasing Probs



VIII Pre-warning Model Evaluation-Part1



Prewarning Precision

P(Fault A Happens|Warning Fault A)

Prewarning Precision stands for the reliability of an engine fault warning.

Increases as thresholds improve.

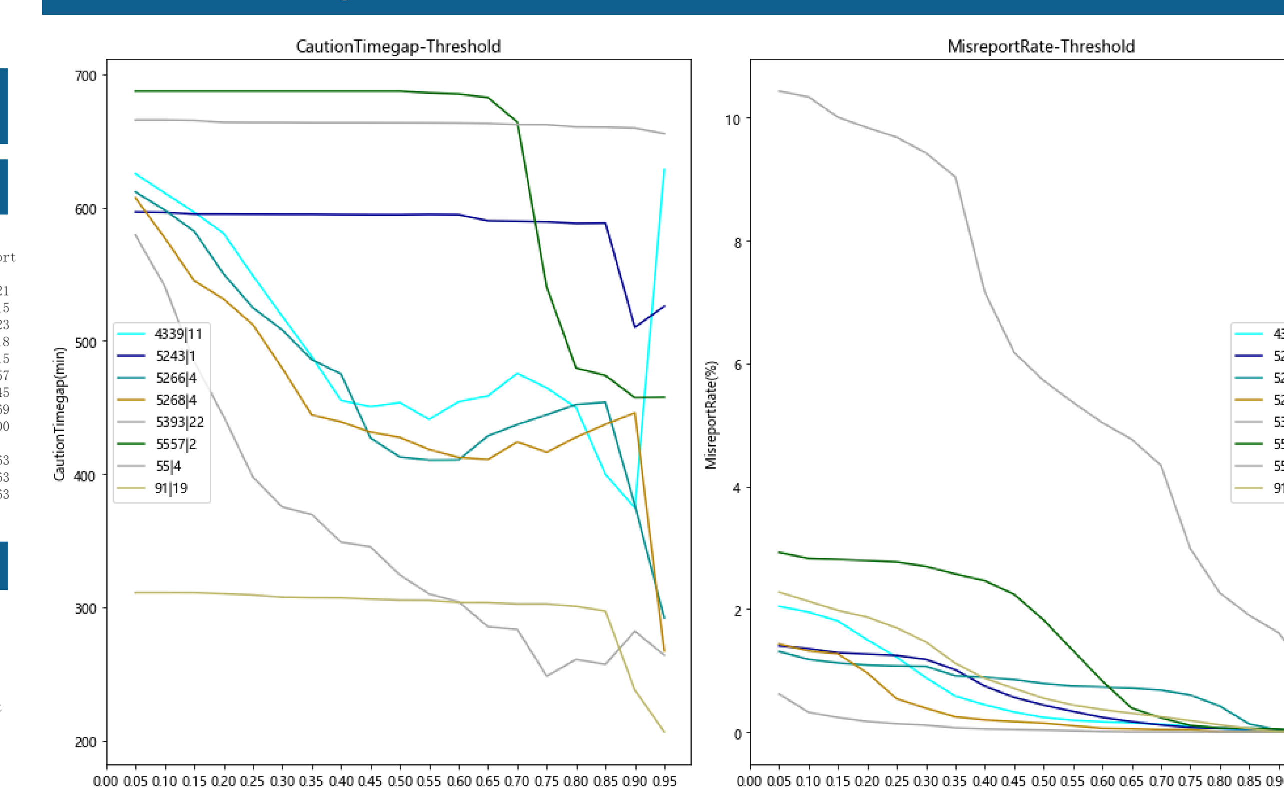
Prewarning Recall

P(Warning Fault A|Fault A Happens)

Prewarning Recall stands for the sensitivity for a type of engine fault warning.

Decreases as thresholds improve.

IX Pre-warning Model Evaluation-Part2



Prewarning Timegap

Difference between timepoint when fault happens and timepoint when a warning is raised.

Prewarning Timegap measures the ability of fault prewarning.

Decreases as thresholds improve.

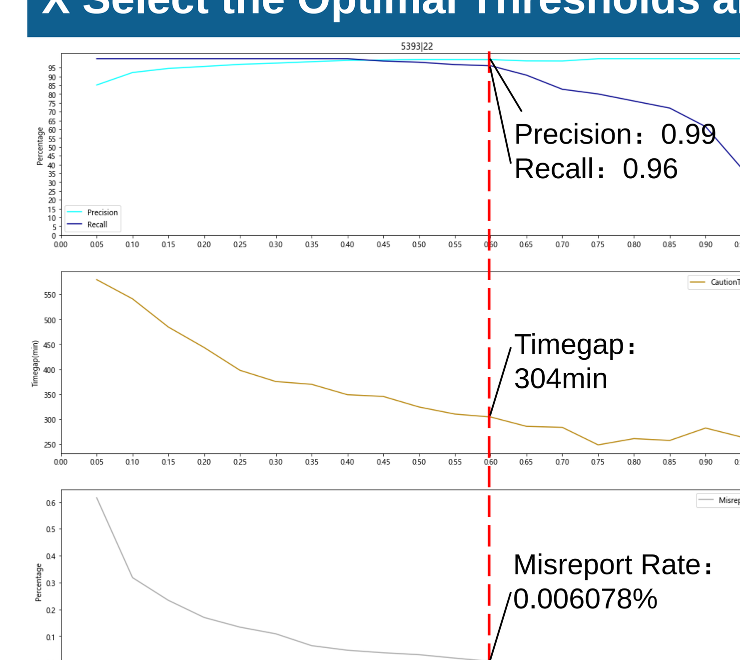
Misreport Rate

The percentage of a false warning.

Misreport Rate is important to enhance user comfort level.

Decreases as thresholds improve.

X Select the Optimal Thresholds and Final Performance



Principles for Threshold Selection

- Precision and Recall come first.
- Increase the timegap as much as possible.
- Lower the misreport rate as much as possible.

Performance	Precision	Recall	Timegap	Misreport Rate
4399|11	0.94	0.94	488min	0.881%
5243|1	0.96	0.96	525min	0.009%
5266|4	0.91	0.78	453min	0.129%
5268|4	0.98	0.90	418min	0.099%
5393|22	0.99	0.96	304min	0.006%
5557|2	0.72	0.95	467min	0.024%
55|4	0.71	0.99	655min	0.983%
91|9	0.99	0.98	206min	<0.001%
Macro Average	0.90	0.93	440min	0.266%