ADSP

Politecnico di Torino

CORN YIELD FORECAST

Parsa Taati

Hamed Goldoust

Afsoon Abbasi

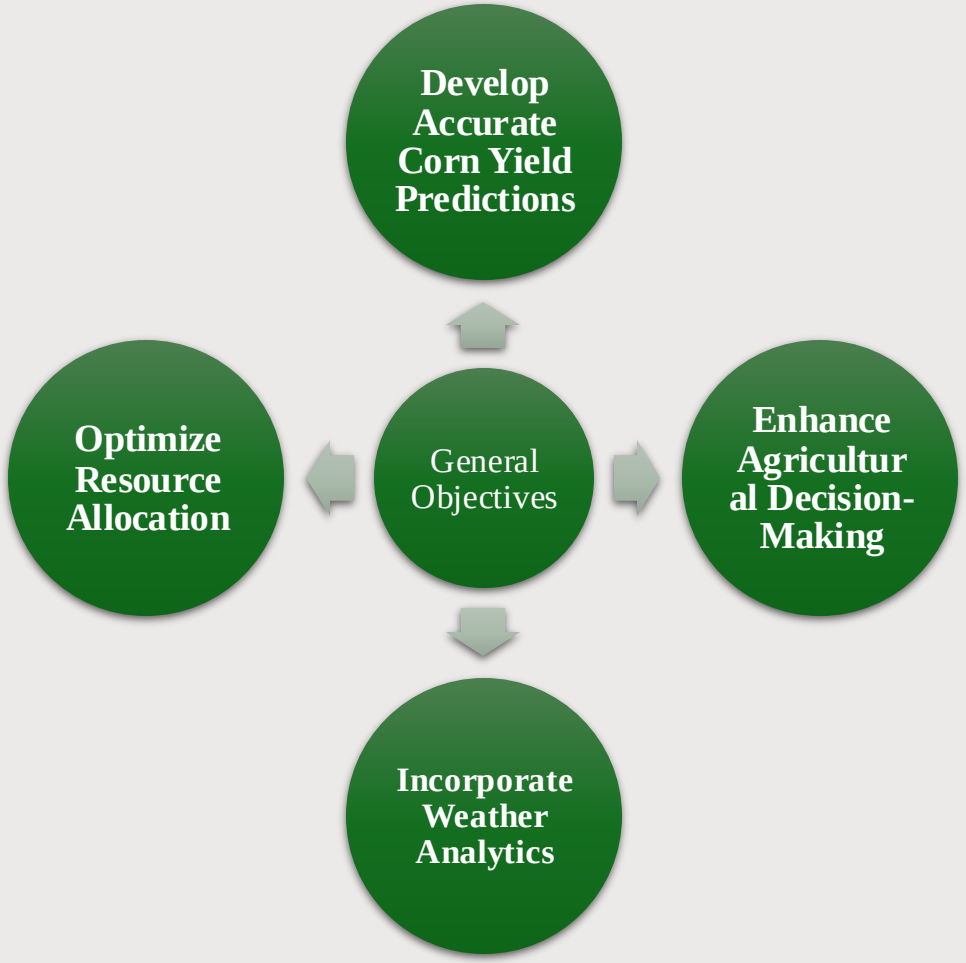
FONDAZIONE links

PASSION FOR INNOVATION

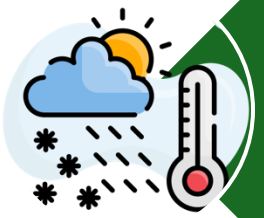
Project Value Proposition

For farmers, internal team members, and academic advisors,
who require actionable insights for crop management, reliable datasets for accurate model development, and well-documented progress for academic evaluation,
our project, Corn Yield Forecast, is a data-driven decision-support system leveraging predictive models and weather analytics,
that provides optimal planting and harvesting recommendations, clean and accessible datasets, and comprehensive project documentation to enhance agricultural efficiency, streamline development processes, and ensure academic rigor.
Unlike existing solutions, our system integrates advanced predictive analytics with user-centric design, ensuring accurate yield forecasting, seamless data accessibility, and actionable insights tailored to the unique needs of farmers, researchers, and development teams, while promoting sustainability and academic excellence.

General objectives



Hypotheses



Higher cumulative rainfall positively correlates with corn yield

Solar radiation significantly impacts crop growth

Solar radiation significantly impacts crop growth

Datasets

Corn_Yield

Targets:

Fresh Total

Fresh Ear Percentage

Fresh Foliage Percentage

Dry Total

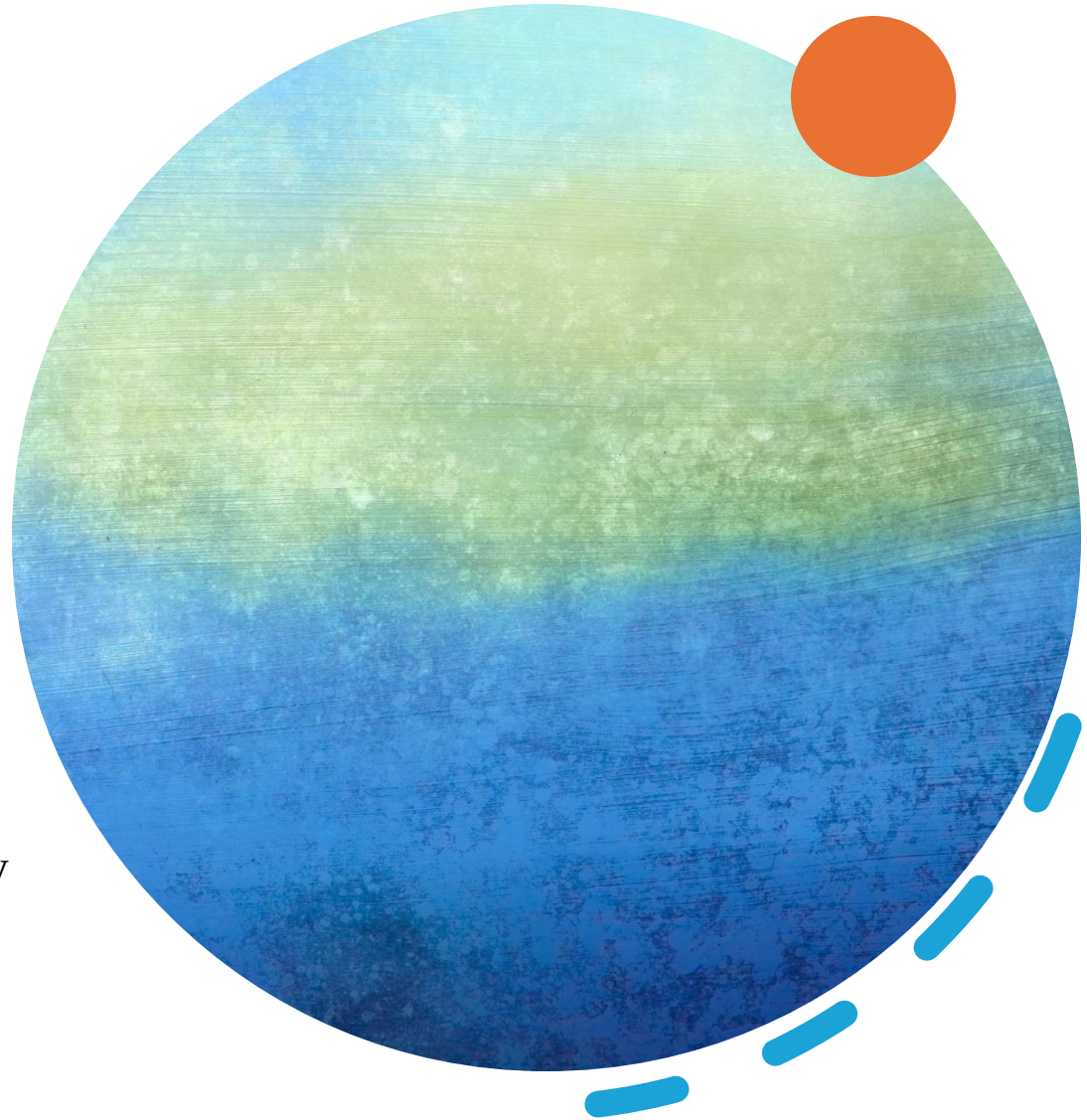
Dry Ear Percentage

Dry Foliage Percentage

Meteo_station

Features:

- wind speed
- maximum soil temperature
- minimum relative humidity
- maximum temperature
- …



We are doing

Step 01

- Data Merging

Step 02

- Time Feature Transformation (sin-cos)
- Label Encoding

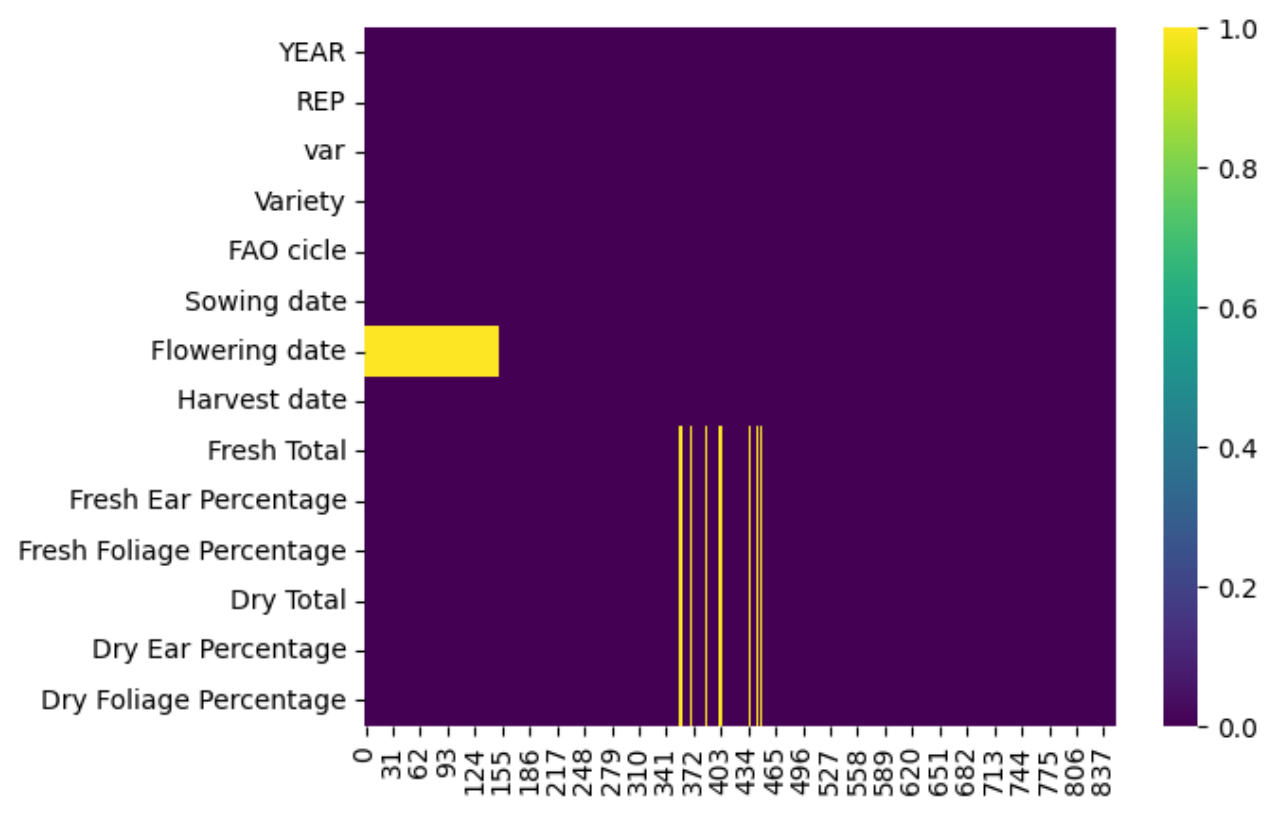
Step 03

- Removing repetitive
- Irrelevant Columns

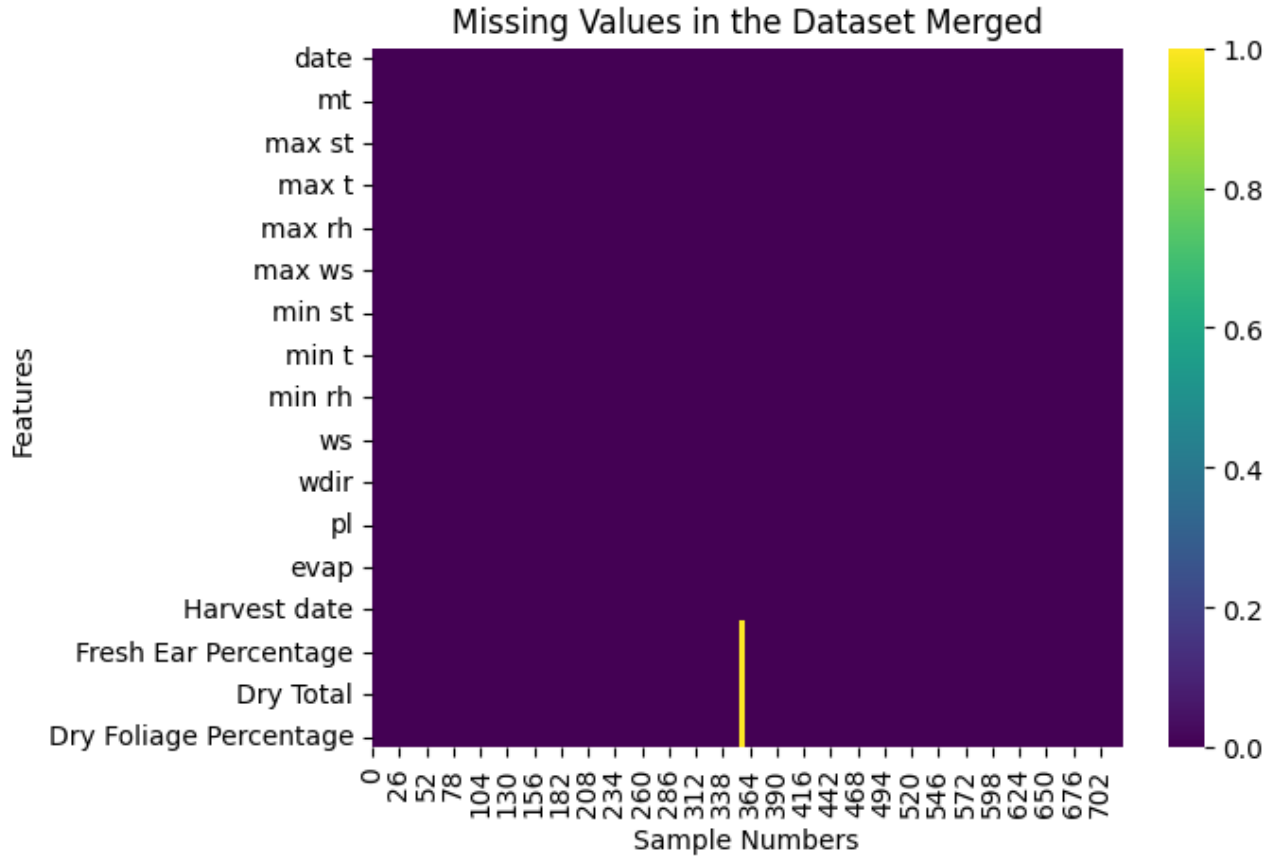
Step 04

- Outlier Detection:Z-scores, IQR Scaling and Normalization: min &max

- Feature Selection:correlation matrix



These are our missing values in “Corn_Yield” dataset before merging



Our missing values after merging

Random Forest Pipeline for Corn Yield Prediction:

Raw Data

Preprocessing

- Data cleaning
- handling missing values
- encoding categorical

Feature Selection

- Analyzing correlations
- selecting the most relevant features

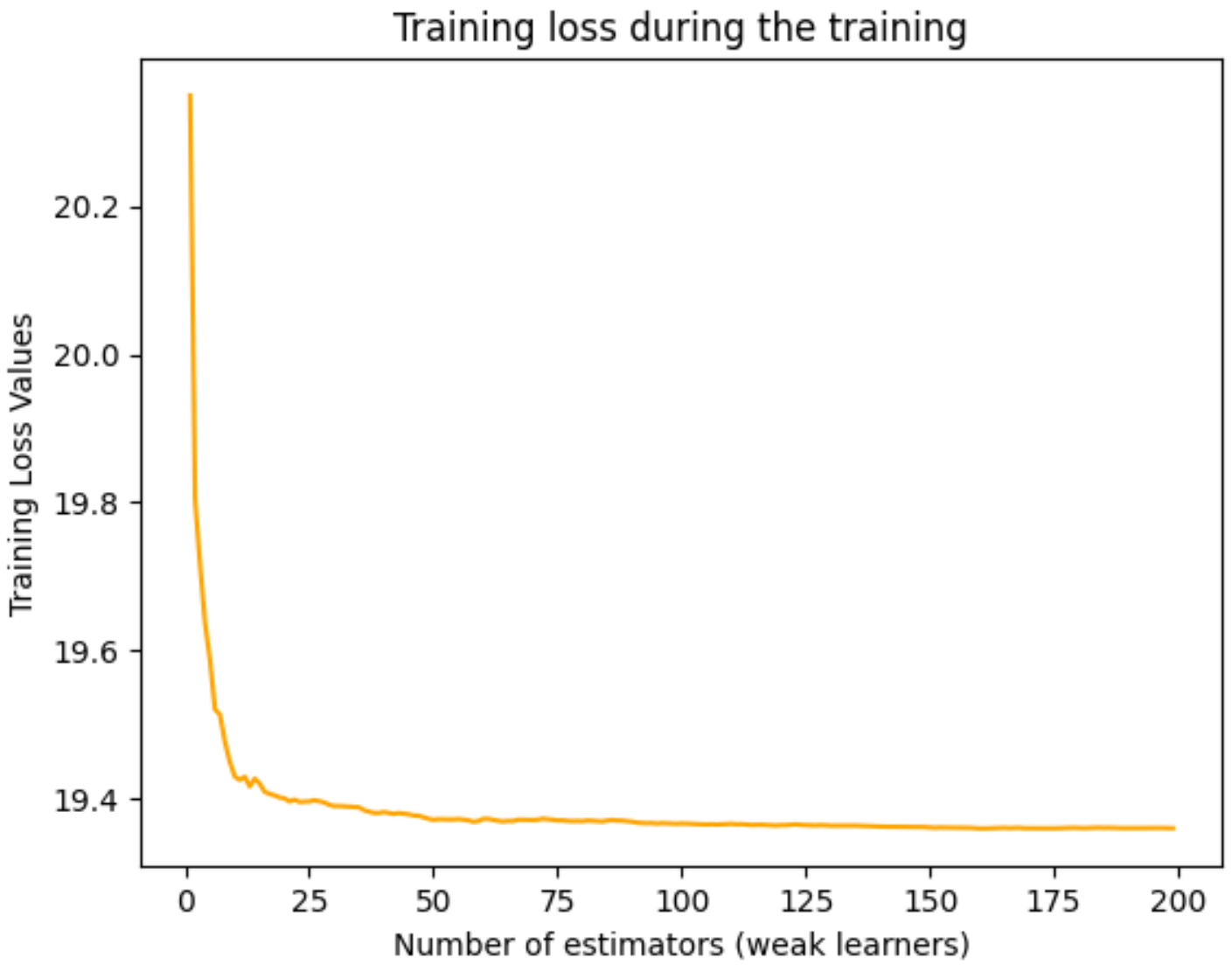
Model Training

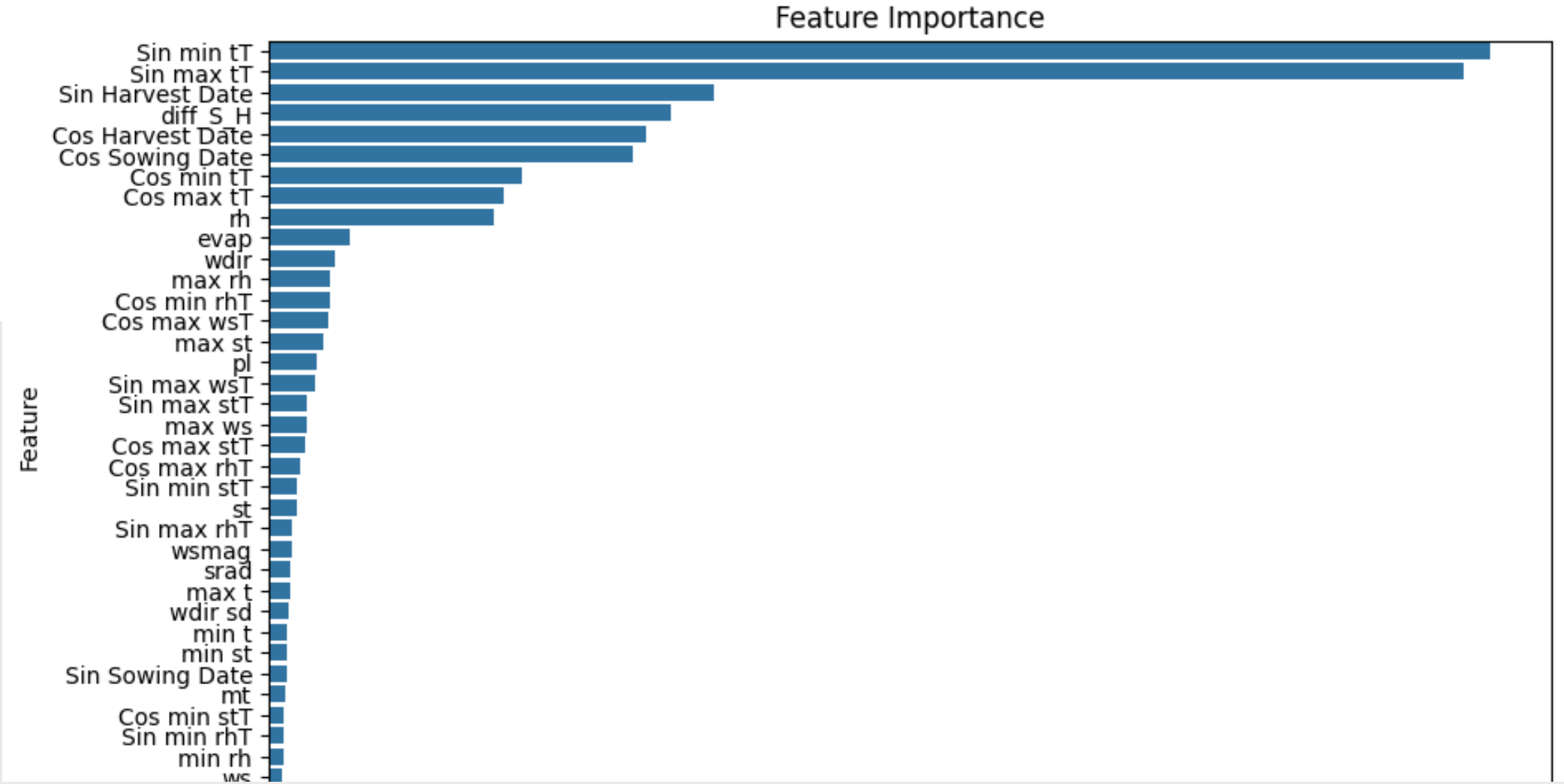
- Training the Random Forest model
- Logistic Regression

Evaluation

- Assessing model performance
- generating yield predictions

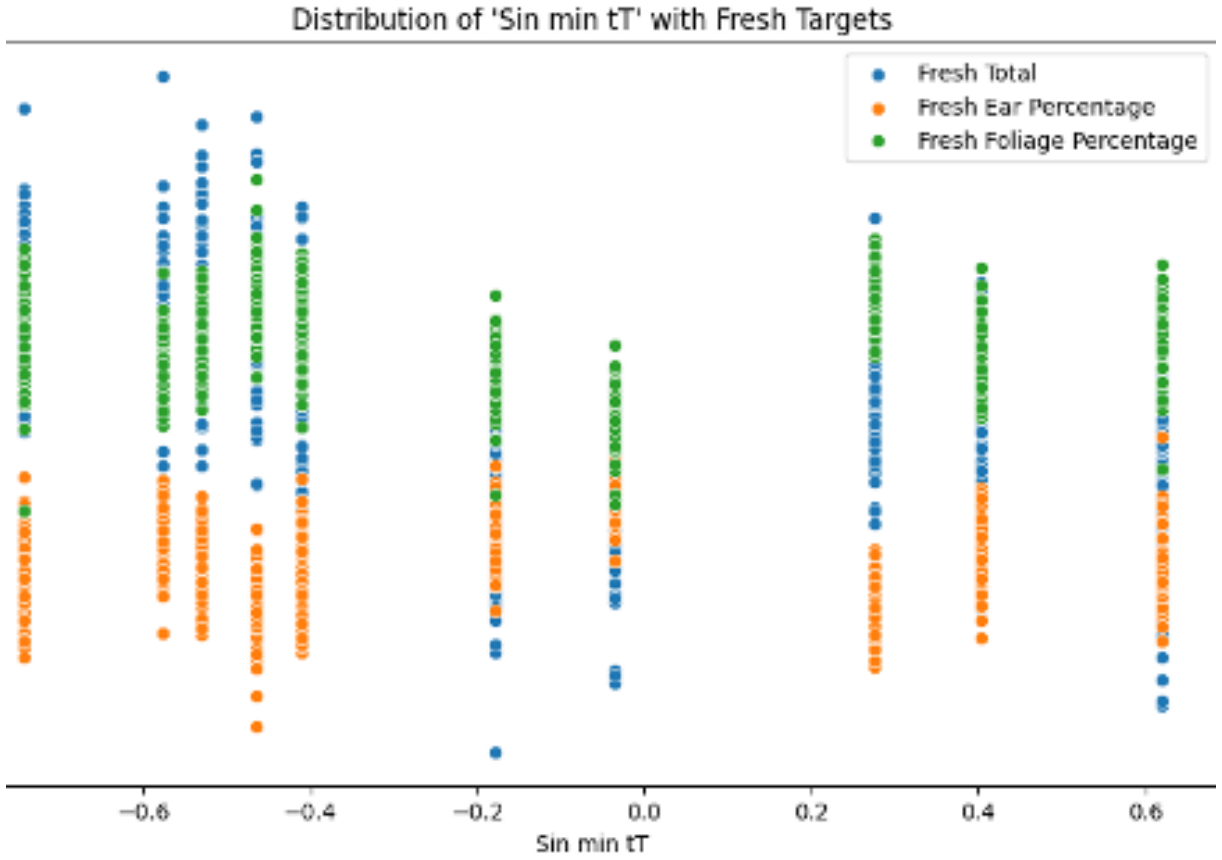
Training Loss

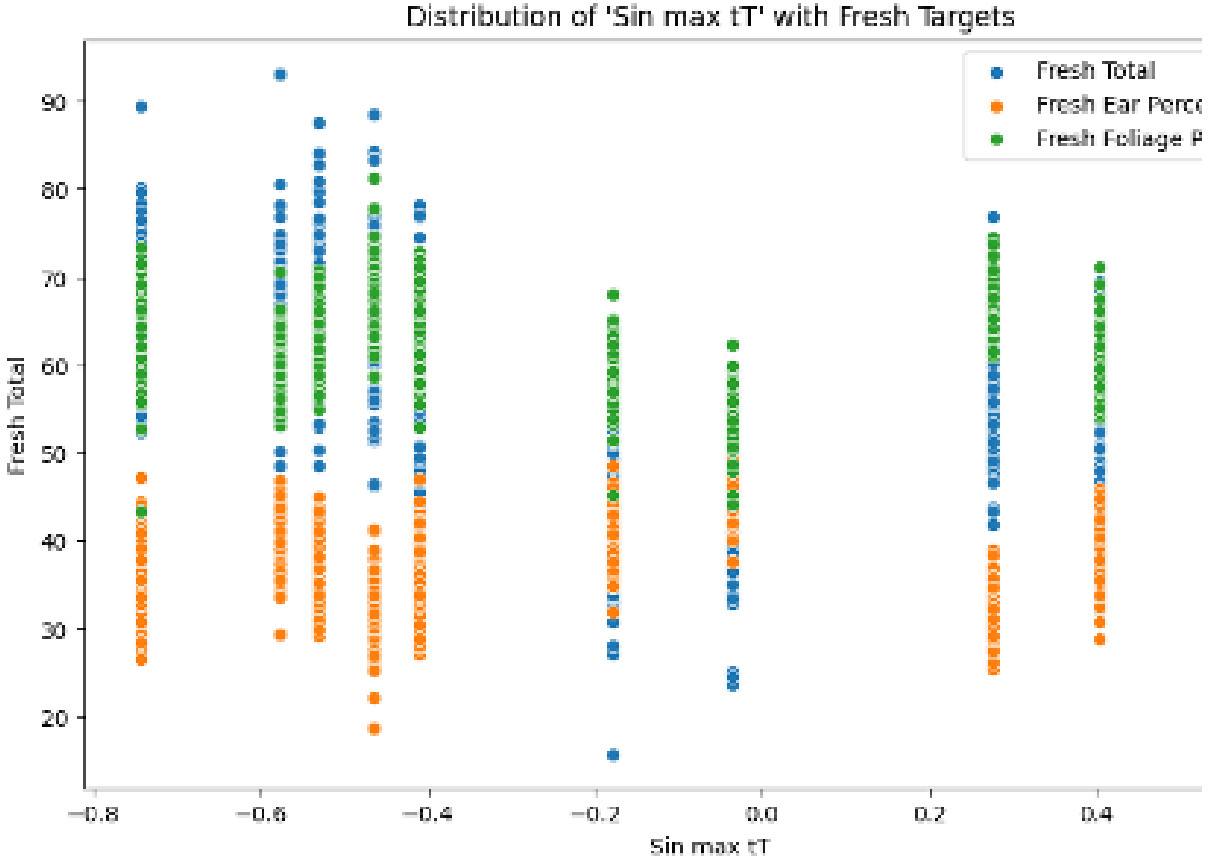




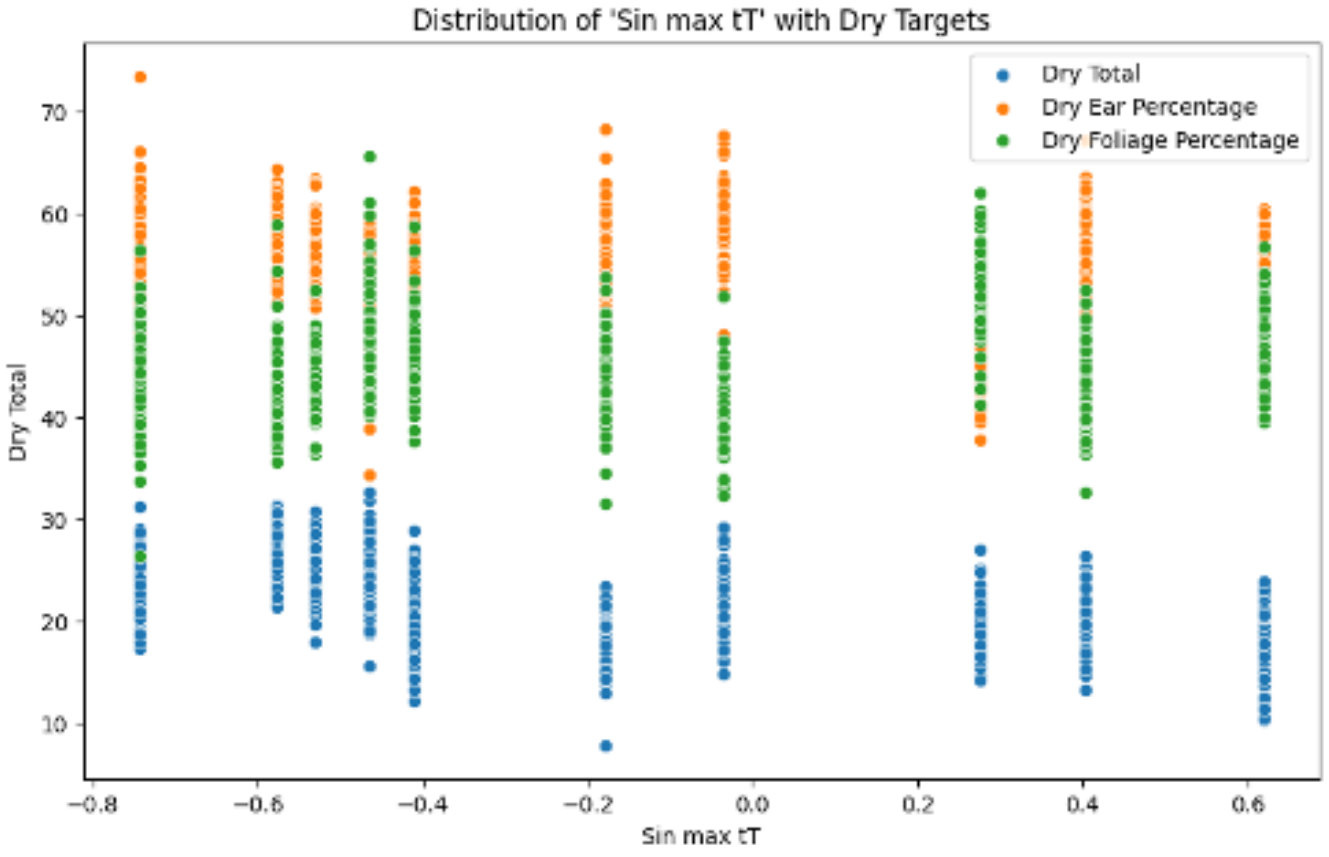
Feature Importance property in Random Forest

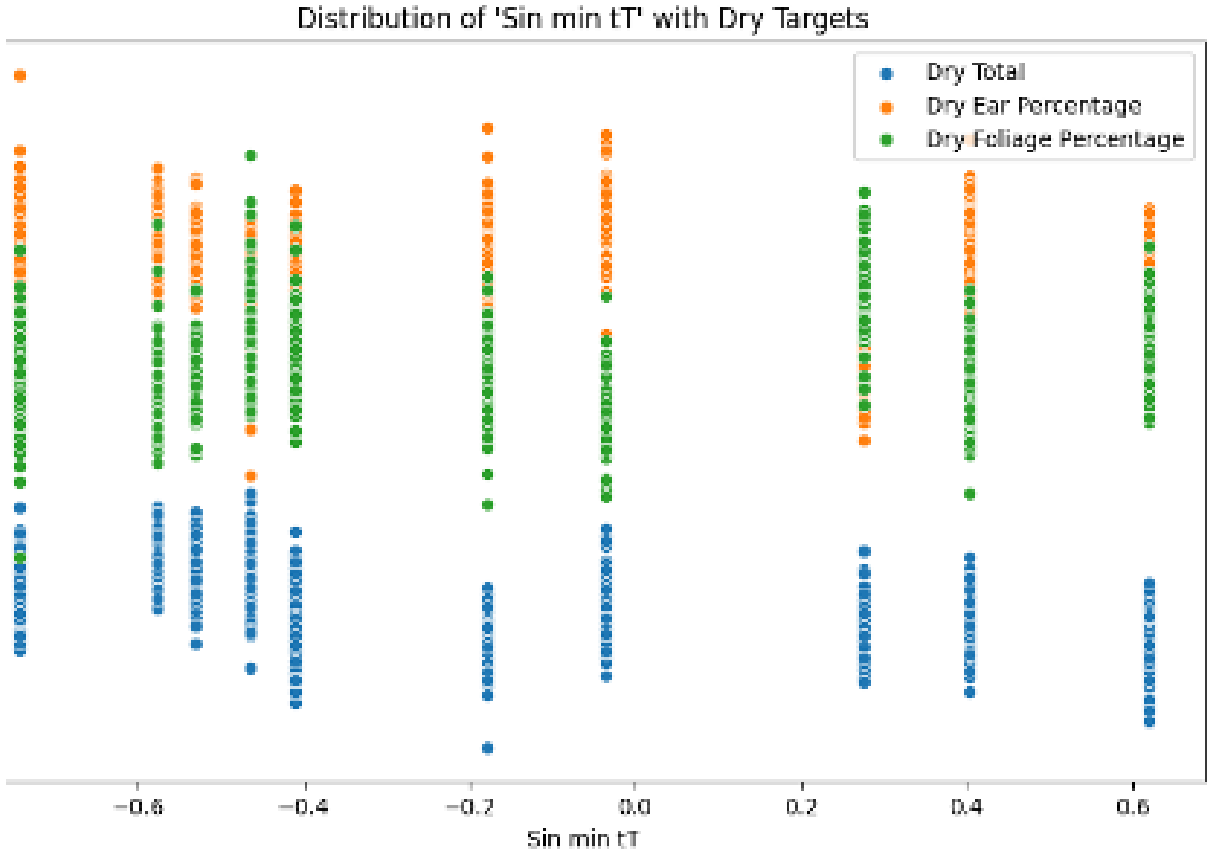
Time Temperature distribution (Fresh Target)



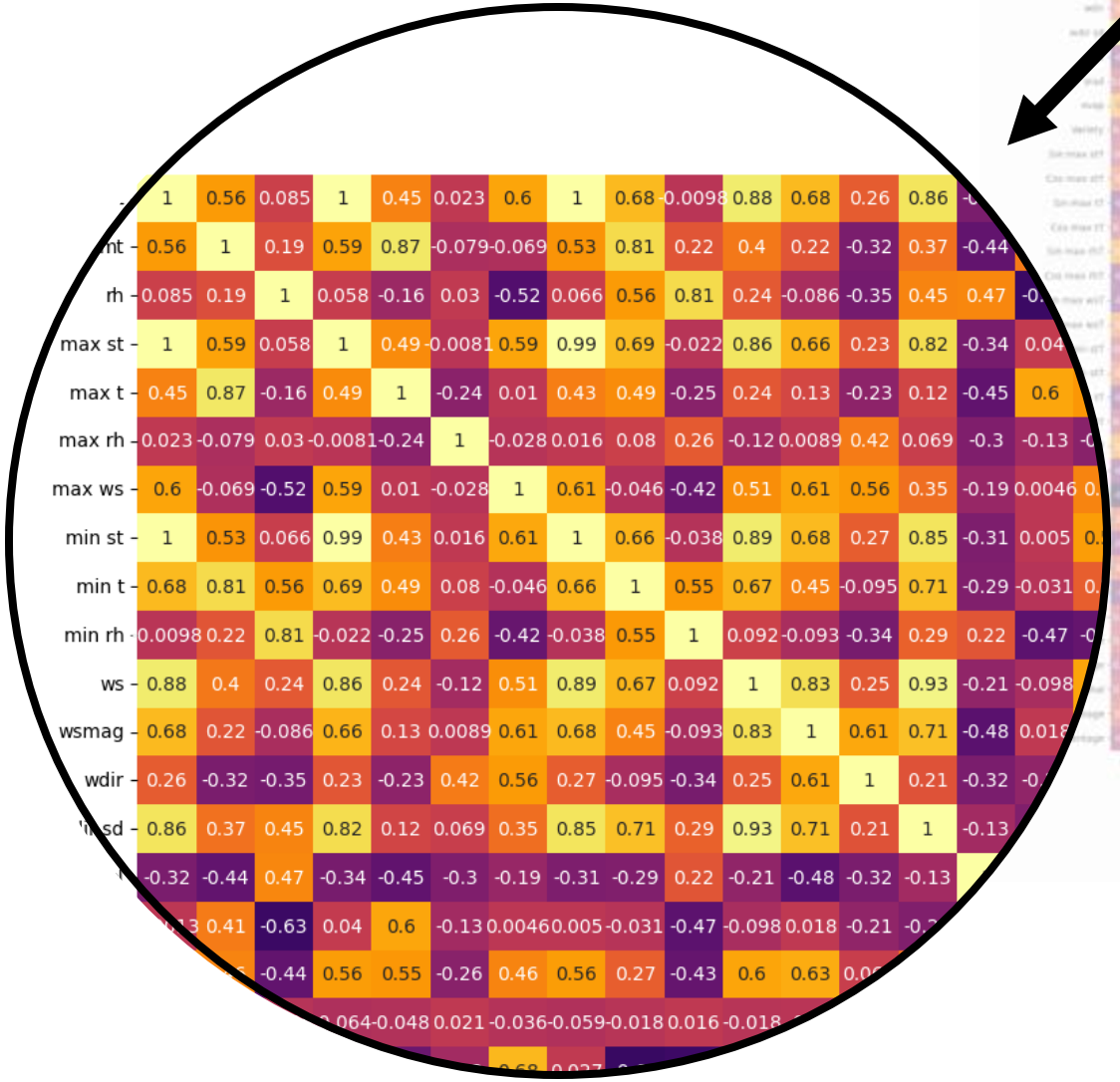


Time Temperature distribution (Dry Target)





- **Correlation Matrix merged dataset**



Result

- **Random Forest results:**
  - Mean Squared Error: 17.534605962105932
  - Mean Absolute Error: 3.0163802001067346

- **Future Develops Model:**
  - Lag features
  - We will work on Deep Learning (GRU)

Thanks for your attention