**Oral presentation script**

Hello, everyone. My name is Wang Zizheng. I'm from Central South University, majoring in Statistics.

Today, I‘d like to present my GEARS research project about **Making Sustainable Development Prediction to Prioritize Sustainable Development Goals (SDGs) Using TabNet.**

**This project aims to build a machine learning model to predict sustainable development given some indicators and prioritize the SDGs based on their importance for sustainable development.**

**The dataset I used is the global indicator framework for the SDGs ~~proposed by the UN~~. It contains more than 200 indicators for over 29 years and ~~more than~~ 200 countries. We also added another sustainable development index dataset to establish supervised labels.**

**This flowchart displays the data preprocessing pipeline.**

**Step one, remove f**eatures with over 80% censoring rate**and highly correlated features.**

**Step two, missing value imputation using the MissForest algorithm.**

**Step three, min-max scaler to scale features into the same range.**

Then, we get a prepared dataset for machine learning.

And let's move on to the model part. The first algorithm that came into my mind for this project is this TabNet model. TabNet is a high-performance and interpretable ~~network~~ architecture designed for ~~learning~~ tabular data. It uses sequential attention to create a mask layer for an instance-wise feature selection, and it can aggregate the instance-wise feature importance to global feature attributes.

These two properties will make this model more advantageous compared to other traditional machine learning models. Because, in this project, different countries put different efforts into achieving the same goal, in this case, the instance-wise feature selection is a better fit for simulating this phenomenon.

To demonstrate TabNet's solid performance, we also tested other popular algorithms. We can see from this table, TabNet outperforms all the others in evaluation metrics such as balanced accuracy and f1 score. And there is room to improve its performance through hyperparameter tuning.

Also, this figure below shows the top 5 most important indicators according to their computed global feature attributes,

We've discussed many points today. Let me ~~quickly~~ summarize the contributions of our project.

First, we built an instance-wise TabNet model to classify sustainable development level ~~based on existing framework~~. Policymakers can put their data into this model and get a predicted sustainable level.

Second, we obtained the five most important indicators for sustainable development. They are **Child mortality rate, Services, value added per worker, Access to electricity, CO2 emissions and GNI per capita,** Countries and regions should adjust their SDG priorities and pace accordingly.

That's all for my presentation. Thank you for your listening.