

ISTANBUL TECHNICAL UNIVERSITY
FACULTY OF COMPUTER AND INFORMATICS ENGINEERING
BLG527E - MACHINE LEARNING
FALL 2025-2026
TERM PROJECT

DUE: JANUARY 5, 2026, 12:00

INSTRUCTORS: Ayse Tosun (tosunay@itu.edu.tr) Yusuf Yaslan (yyaslan@itu.edu.tr)

RULES:

1. Work in groups of **two**. Build your own team; if you cannot find a team, you are allowed to implement your project on your own.
2. Submit the following to **Ninova** by the deadline: Your term project report in [ACM Conference Proceedings](#) format, five-minutes demo, and source codes written in Python.
3. Submit your solution(s) to the competition website: [DengAI](#)
4. Late submissions will NOT be accepted. Submissions via e-mail will NOT be accepted.

You are going to **join a competition** during this term project. The competition is released by DrivenData (<https://www.drivendata.org/>) and it is currently open for submissions. The objective is to **predict the number of dengue fever cases each week** (in each location) based on environmental variables describing changes in temperature, precipitation, vegetation, and more. The competition has been open for a while, since the objective is to promote learning and exploring. You may find the leadership board with the **Mean Absolute Error** rates reported by each submission. The dataset is available once you join the competition.

For this term project, you are expected to **propose a model** that would make as good predictions as possible. You may **check online repositories or solutions** published by others, and **cite those** during the development of your work. There have been several, relatively recent **publications** on dengue prediction in the literature (e.g. 1-3). Choose **at least two algorithms, one being a polynomial of a high (k) degree** – choose k according to the data– and the **other being utilized or proposed in these prior works**. **Compare** and assess the algorithms in terms of the prediction **error**, as well as in terms of how well the algorithms utilize the dataset characteristics. Report which characteristics are the best indicators of the output variable.

This competition aims to predict the dengue cases each week. Your initially proposed model would make predictions each week, but as you have a time series data, your second task during this term project will be to predict the dengue cases beforehand. **For this aim, propose a second model** that would make **predictions a couple of weeks before**; and report what is the best period to predict the dengue cases beforehand.

Do not forget to submit your best performing solution to the competition. Solutions with no submission to the competition will not be evaluated.

Sample previous works

1. Md. Siddikur Rahman, Md. Abu Bokkor Shiddik, Explainable artificial intelligence for predicting dengue outbreaks in Bangladesh using eco-climatic triggers, Global Epidemiology, Volume 10, 2025, <https://doi.org/10.1016/j.gloepi.2025.100210>.
2. Madewell, Z.J., Rodriguez, D.M., Thayer, M.B. et al. Machine learning for predicting severe dengue in Puerto Rico. Infect Dis Poverty 14, 5 (2025). <https://doi.org/10.1186/s40249-025-01273-0>.
3. Sebastianelli, A., Spiller, D., Carmo, R. et al. A reproducible ensemble machine learning approach to forecast dengue outbreaks. Sci Rep 14, 3807 (2024). <https://doi.org/10.1038/s41598-024-52796-9>.