

Benchmark Performance

A minimum expectation of performance from a prediction system or machine learning model.



Methods to Obtain Benchmark Performance

- Human Performance
- Performance of an earlier prediction system
- Performance that will make a prediction system useful



Human Performance

The performance of humans on a particular task can be considered a benchmark for prediction systems in some cases.



Our Model:
Prediction System to
Predict Pneumonia

Expected Accuracy \geq 93%



Benchmark Model:
Performance of Doctors
in Pneumonia Prediction

Accuracy: 93%

Human Performance

The performance of humans on a particular task can be considered a benchmark for prediction systems in some cases.



Our Model:

Classification model that predicts products that will sell > 1000 units

Expected F1-Score ≥ 0.8



Benchmark Model:

Predicts products that will sell > 1000 units

F1-Score: 0.8



Performance of an Earlier Prediction System

Performance of an Earlier Prediction System

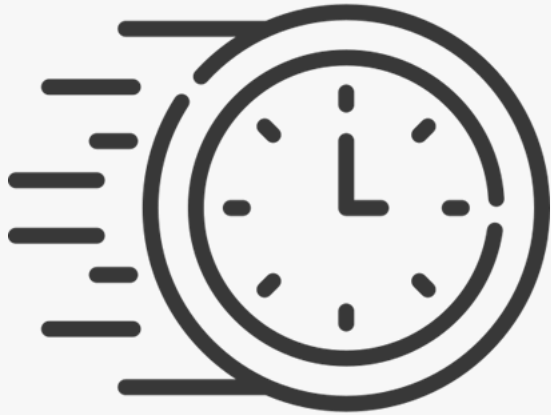
The existing prediction system as a benchmark can be used in two scenarios:

- When the prediction system is already better than humans
- When the prediction system is not as good as humans
but is more efficient



Performance of an Earlier Prediction System

- When the prediction system is already better than humans



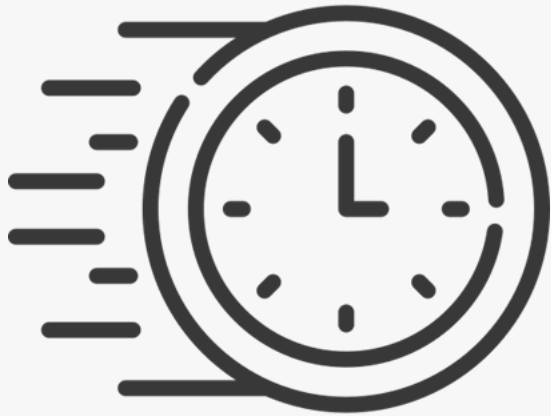
Expected Time of Arrival
at a destination



Google Maps

Performance of an Earlier Prediction System

- When the prediction system is already better than humans



Expected Time of Arrival
at a destination

- Collecting more data
- Adding more meaningful variables
- Training newer machine learning algorithms

Performance of an Earlier Prediction System

- When the prediction system is not as good as humans but is more efficient



Language Translation

Performance of an Earlier Prediction System

- When the prediction system is not as good as humans but is more efficient

Human Translator

- 1) Expensive to hire
- 2) Fatigue while working
- 3) Translates 3 to 4 languages

Machine Learning Translator

- 1) Not as good as a human translator
- 2) Cheaper and does not face fatigue
- 3) Can translate hundreds of languages



Performance which will make
a prediction system useful

Performance which will make a prediction system useful

Certain scenarios that are predicted for the first time do not have a human reference or existing prediction systems.

- Stakeholders working on the task should fix the benchmark



Performance which will make a prediction system useful (Synergix)



Regression Benchmark:

For predicting number of units sold is **R^2 Score ≥ 0.7**

Benchmarks for Synergix



Classification Benchmark:

F1-Score ≥ 0.8

Regression Benchmark:

R^2 Score ≥ 0.7