Research Prediction Competition

Web Traffic Time Series Forecasting

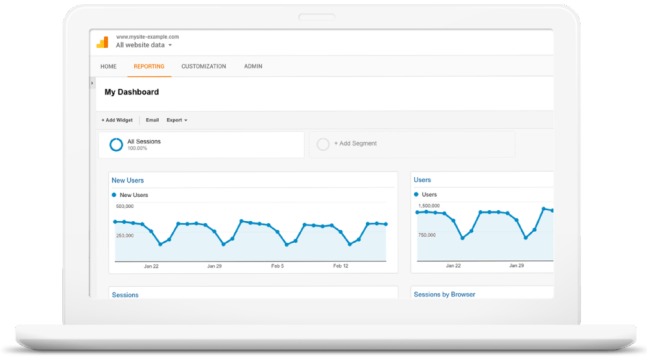
Forecast future traffic to Wikipedia pages

$25,000
Prize Money

Google · 112 teams · 2 months to go (2 months to go until merger deadline)

Overview

This competition focuses on the problem of forecasting the future values of multiple time series, as it has always been one of the most challenging problems in the field. More specifically, we aim the competition at testing state-of-the-art methods designed by the participants, on the problem of forecasting future web traffic for approximately 145,000 Wikipedia articles.



Sequential or temporal observations emerge in many key real-world problems, ranging from biological data, financial markets, weather forecasting, to audio and video processing. The field of time series encapsulates many different problems, ranging from analysis and inference to classification and forecast. What can you do to help predict future views?

This competition will run as two stages and involves prediction of actual future events. There will be a training stage during which the leaderboard is based on historical data, followed by a stage where participants are scored on real future events.

You have complete freedom in how to produce your forecasts: e.g. use of univariate vs multi-variate models, use of metadata (article identifier), hierarchical time series modeling (for different types of traffic), data augmentation (e.g. using Google Trends data to extend the dataset), anomaly and outlier detection and cleaning, different strategies for missing value imputation, and many more types of approaches.

We thank Google Inc. and Voleon for sponsorship of this competition, and Oren Anava and Vitaly Kuznetsov for organizing it.

Kaggle is excited to partner with research groups to push forward the frontier of machine learning. Research competitions make use of Kaggle's platform and experience, but are largely organized by the research group's data science team. Any questions or concerns regarding the competition data, quality, or topic will be addressed by them.

Leaderboard

1 CPMP
2 Nikita Mikhaylov
3 Morgrey
4 Pan Tofelek
5 Bojan Tunguz
6 SergazyKalmurzayev
7 Jason Benner
8 Oleg Panichev

Kernels

Wiki Traffic Forecast Explor...
12 votes · 3 hours ago

Simple Model
11 votes · 8 hours ago

R benchmarking SMAPE fu...
2 votes · 8 hours ago

8 weeks median
6 votes · 5 hours ago

Preliminary Investigation: H...
2 votes · 5 hours ago

26 discussion topics

Share two Python libraries t...
0 replies · 5 hours ago

SMAPE - R
6 replies · 6 hours ago

Machine configuration
2 replies · 7 hours ago

SMAPE for R with C++ roun...
0 replies · 8 hours ago

SMAPE_Python
3 replies · 9 hours ago

Launch
3 days ago

Close
2 months

2 months
Rules Acceptance Deadline

112 Teams **112** Competitors

Points **This competition awards standard ranking points**
Tiers **This competition counts towards tiers**