

Она призвана справляться с пробелами в данных или определять ситуации более точно, чем существующие алгоритмы. Модель изучает причинно-следственные связи на основе синтетических данных, что делает ее более надежной, особенно для небольших наборов данных с менее чем 10 000 строк, провалов или большого количества пропущенных значений.

TabPFN была обучена на 100 миллионах синтетических наборов данных, что позволило научить ее оценивать различные причинно-следственные связи и использовать их для предсказаний. В этих случаях она превосходит такие алгоритмы, как XGBoost, требуя лишь половину данных для достижения той же точности, что и традиционные модели.

Модель также способна адаптироваться к новым типам данных, подобно тому, как обучаются большие языковые модели.