Predibase

Customer Story:

checkr

Streamlining background checks with small language models

About Checkr

Checkr is a technology company that specializes in delivering modern and compliant background checks for over 100,000 customers, including Warby Parker, Uber, Dominos, Doordash, and Kimpton Hotels. As a platform that combines the best in human expertise, AI, and machine learning, Checkr has fairness and a focus on the candidate experience embedded into the platform. Checkr makes it easy for people to understand reports and make fair decisions and enhances the efficiency, inclusivity, and transparency of the background check process.

Checkr provides an automated solution for adjudication, which is the process of reviewing background check results to determine if a candidate is suitable for hiring based on a company's policies. Checkr's automated solution removes the need for manual review by 95%. To do so, Checkr collects data from various vendors to analyze candidates, creating a process that is unbiased, transparent, and compliant.

The Challenge

Checkr conducts millions of background checks each month, processing thousands of rows of unstructured text data for each check. On average, 98% of the data is processed efficiently using a proprietary ML model. The remaining 2%, however, are highly complex datasets that require categorizing data into over 200 distinct categories for customers to adjudicate their hires automatically.

These complex adjudication cases require careful handling and have a number of challenges and requirements:

- **COMPLEX DATA**
  2% of categorization tasks are complex, involving very noisy unstructured text data that is too challenging for Checkr's ML model to accurately process. This results in time-consuming manual human review of the data. They wanted to use LLMs to address this challenge.
- **REAL-TIME INSIGHTS**
  The adjudication process is made up of synchronous tasks, meaning the tasks must be completed in order, but must meet low-latency SLAs in order to provide Checkr's customers with near real-time employment insights.
- **HIGH ACCURACY REQUIREMENTS**
  Checkr's reports are used to make important decisions about a prospective employee's future. Therefore, it's critical that models are highly accurate.
- **BALANCING COST**
  Achieving Checkr's efficiency and accuracy goals must be balanced with maintaining reasonable inference costs.

The Solution

As an alternative to their traditional ML models, Checkr explored several approaches to improve classification accuracy with LLMs and several fine-tuning and inference platforms, evaluating on accuracy, efficiency, cost-effectiveness, and ease of use. After this evaluation, Checkr chose Predibase due to its ability to consistently provide highly accurate results. Leveraging a small open source LLM, llama-3-8b-instruct, and fine-tuning it on Predibase, **Checkr was able to achieve an accuracy of over 90%** for the most challenging 2% of cases, outperforming both GPT-4 and all other fine-tuning platforms.

Because Predibase is an end-to-end platform for the entire generative AI lifecycle — from model training and deployment to ongoing optimization in production — Checkr was able to fine-tune and serve high-performing small language models that outperform GPT-4 without needing to manage complex infrastructure. Beyond accuracy, they saw a number of benefits from working with Predibase including:

LOW-LATENCY INFERENCE

Predibase is specifically optimized to deliver low-latency for fine-tined inference and was able to consistently deliver 0.15 second response times for Checkr's production traffic, which is 30x faster than the GPT-4 experiments that Checkr ran. This low-latency was key in enabling Checkr to meet the needs of their customers.

SIGNIFICANT COST SAVINGS

By fine-tuning and serving Small Language Models (SMLs) on Predibase, Checkr reduced inference costs by 5x compared to GPT-4. Predibase enabled Checkr to expand on multiple use cases using multi-LoRA serving on LoRAX, which will further bring costs down as they scale with more use cases. These cost savings enable Checkr to keep cost low for their customers.

EASY-TO-USE PLATFORM

Beyond a robust and developer-friendly software development kit (SDK), Predibase provides Checkr with a user-friendly web app which makes it easy to manage projects, track model versions, and explore model performance with visualizations. Predibase's dashboards for production metrics help teams at Checkr understand production efficiency and performance over time

By working with Predibase to improve classification accuracy and streamline background checks at scale, Checkr achieved:

90%+
accuracy
for their classification tasks, outperforming GPT-4

30x
faster response
serving their fine-tuned SLMs in production

5x
cost reduction
using Predibase vs GPT-4

About Predibase

At Predibase, we are committed to making AI accessible for everyone. Our platform streamlines the process of training and deploying small language models, allowing teams to quickly harness the power of their data to build and deploy highly accurate GenAI applications at a fraction of the cost. With a focus on user experience, we empower organizations to collaborate and innovate effectively, overcoming complex AI challenges and delivering actionable insights back to their business.

To learn how Predibase can help your teams deliver GenAI value faster:

Schedule a custom demo with an LLM expert:
https://pbase.ai/demo

Get started fine-tuning and serving SLMs for free:
https://pbase.ai/getstarted

> "At Checkr, we have been getting first class support from Predibase. Because it is built on open source software, the platform and the company are extremely transparent. The Predibase team does a great job of continuously educating us, being very user friendly and building trust by being at the apex of their craft."

VLAD BUKHIN
STAFF ML ENGINEER AT CHECKR