Pastel Reduces Data Preparation Time by up to 80%

June 2024 | Case Study

80%

Saving time

in data preparation of your NLP project lifecycle time allocation.

01 The Story of Pastel

Pastel's mission is to be the leading AI enterprise solutions provider for Africa, beginning with specialized financial modules and ultimately expanding to a variety of other enterprise applications. They're looking to pioneer AI-driven solutions tailored to the African context, ensuring businesses across the continent access world-class intelligence for optimized operations. Their commitment to innovation and user-centric approaches ensures that they continuously adapt to the needs of their users, making financial management more accessible and efficient for all.

02 The Challenge of Manual Data Labeling

With anyone at the forefront of innovation, Pastel faced significant challenges in managing extensive NLP labeling tasks manually. This manual process led to inefficiencies and slowed down project timelines, necessitating a more streamlined and automated approach.

Pastel deals with vast amounts of financial data that require precise and granular labeling. For example, customer requests with banking institutes need to be categorized accurately to train AI models effectively. The manual labeling process was time-consuming and prone to errors, which hindered the efficiency and scalability of their AI projects.

03 Why Pastel Chose Datasaur

Pastel needed a robust tool that could handle complex and layered data labeling efficiently. In other words, they needed to automate their labeling process.

Too much time and resources were being spent on the manual labeling process. Their manual method for labeling lacked automation features, token-search functionalities, and was not user-friendly.

Datasaur offered a comprehensive solution tailored to handle their complex NLP labeling projects with ease. With Datasaur, Pastel could integrate various LLM base models, deploy ML-Assisted Labeling, and utilize a robust QA tooling system to evaluate and ensure the quality of labeled data.

05 How Datasaur Responded

Automation

Datasaur offered Pastel three tools specifically for their workflow; these tools automated the labeling process in their workflow.

- **ML-Assisted Labeling**: Enabled direct integration to HuggingFace, spaCy, and Pastel's own custom models to automatically return labels.
- **LLM Labs**: Through LLM Labs, Pastel is able to compare over 120 base LLMs, fine-tune a model, and then deploy that model to automatically label their data.
- **Predictive Labeling**: Enabled Past Africa's team to label only a few entities, and the Datasaur platform will actively learn from their labels, and predict the rest of their dataset.

> "Datasaur's automation suite improved the efficiency of our labeling with incredible effectiveness. It saves our team valuable time and resources.
>
> —Pastel

Ease of Use

Pastel was impressed with Datasaur's intuitive UI/UX. The platform allowed them to efficiently manage their extensive taxonomy of labels, utilize hotkeys for quick label application, and easily incorporate the automation tools mentioned above.

For example, to deploy Predictive Labeling, the labeler only needs to make two clicks. There is no coding, nothing that requires technical experience or knowledge. The labelers at Pastel were able to apply a few labels, click twice for deploy Predictive Labeling, and then the rest of the dataset is labeled automatically.

> “We found the entire platform incredibly intuitive and easy to navigate. Onboarding was smooth and we were able to quickly adopt their automation tooling –- which was very important for us when considering a labeling platform.
>
> —Pastel’s Project Manager

Pastel is using Datasaur’s publicly hosted site which is hosted on AWS. They are also utilizing Amazon’s S3 data storage to integrate their data with Datasaur. By using their S3 buckets, they can securely transfer data to Datasaur for project creation and exportation. This has created a seamless data pipeline for Pastel – in which the projects are created/exported all securely and automatically within the AWS environment.

Genuine Customer Support

Pastel has found value in the dedicated support from Datasaur's customer success team. They received prompt assistance for any issues, ensuring smooth and uninterrupted project workflows. The support team also set up multiple onboarding sessions to ensure Pastel knew how to use the platform. These onboarding sessions included, but were not limited to, one session dedicated entirely to deploying automation for their labeling process.

06 The Results

After transitioning to Datasaur, Pastel observed several significant improvements, and anticipates being able to reduce labeling time by 80%:

1. **Improved Tracking and Review Processes:**
   With Datasaur, Pastel gained better insight into their labeling projects. They could easily track progress, identify roadblocks, and resolve inter-annotator disagreements swiftly.

2. **Increased Project Efficiency due to Automation:**
   The efficiency of their labeling projects improved drastically. Thanks to the automation suite, the overall labeling process was more streamlined. Datasaur's automation capabilities allowed Pastel to scale their operations and improve the accuracy of their labeled data.

07 Conclusion

By leveraging Datasaur's NLP labeling automation tools, Pastel transformed their data annotation processes, resulting in increased efficiency, accuracy, and scalability. The collaboration with Datasaur enabled Pastel to focus on their core mission of enhancing financial inclusivity and empowering businesses in Africa.

About Datasaur

Datasaur is a private LLM provider and data labeling platform designed for companies to build their AI ecosystem with ease and efficiency. It assists organizations and universities in setting up custom LLMs and annotating data more efficiently and accurately through automation, quality control, and human-in-the-loop workflows. For more information, visit www.datasaur.ai.

Schedule a demo