

Nightfall Achieves 2.5x Labeling Gains and 2.5x Fewer False Positives with Datasaur



2.5x

labeling gains

2.5x

fewer false positives

01 Introduction

Nightfall, after comparing 55 tools, chose Datasaur: gaining new QA functionality. This switch saved time and significantly boosted productivity.

02 Nightfall AI Mission Statement

Nightfall AI's mission is to empower organizations to protect sensitive data across digital environments using advanced, AI-driven security. With a focus on simplicity and effectiveness, Nightfall AI helps security and IT teams monitor, detect, and respond to data risks in real-time across SaaS platforms, GenAI tools, emails, and endpoints. Their AI-native approach not only strengthens compliance but also enhances data leak prevention, allowing organizations to innovate and operate confidently in a secure digital landscape. Through this partnership, Nightfall AI enables customers to achieve comprehensive data protection while simplifying data security management.

03 Problem Statement

Nightfall, operating in the security industry, grappled with rigid limitations in their previous labeling platform. This inflexibility prevented them from iterating or modifying previously labeled data, leading to significant QA issues. Consequently, they were compelled to label large quantities of data to drown out inaccuracies, necessitating the need for more volume. Their only solution for QA issues was volume. This led to an increase in labor costs and lack of confidence in their labeled data. Subsequently, they gathered a list of 55 platforms to compare as their next solution.

04 Proposed Solution & Architecture

To tackle Nightfall's challenges, Datasaur introduced several key features. First, Reviewer Mode streamlined the QA process by quickly identifying disagreements between labelers and automatically accepting labels with enough consensus. Next, the Multi-pass feature enabled multiple labelers to work on the same data, fostering consensus and enhancing data quality.

“ Datasaur’s QA tooling significantly increased our confidence in the quality of our data.


— Nightfall Project Manager

This was the first time in Nightfall's history that they were able to take a multi-pass approach. Hotkey configurations and keyboard shortcuts allowed labelers/reviewers to efficiently label and review data using keyboard shortcuts, driving efficiency.

“ The actual mechanisms of the labeling are very fluid and we’re a point of efficiency gains.


— Nightfall Project Manager

Lastly, the flexibility in editing label spans, including width and the types of spans that can be executed, provided Nightfall with a tailored solution that met their specific needs. Finally, Datasaur’s QA reports enabled the admin of Nightfall to easily trace inaccuracies previously labeled so they could remedy their labeled data – with full transparency.



The Nightfall team was able to utilize all of these features within their AWS ecosystem. Through the uploading and exporting of data through API within their AWS pipeline, they're able to take advantage of Datasaur's features seamlessly. On this ability, they stated:

“ Our integration with AWS has been smooth, enabling us to leverage cloud storage capabilities and the AWS ecosystem for our data labeling projects effectively. This synergy has streamlined our workflows and improved data management.”



05 Outcomes of Project & Success Metrics

In just one year with Datasaur, Nightfall labeled 180,000 documents and uncovered 97,000 instances of sensitive information, such as passport numbers and social security numbers. This resulted in a 2.5x increase in labeling efficiency compared to their previous operations. Additionally, they saw a 2.5x reduction in false positive rates in their models and a 1.5x improvement in model recall, capturing 50% more entities than before. Furthermore, Nightfall achieved a 1.5x increase in the number of models with labeled data driving their output. This surge in productivity is especially impressive considering they previously relied on a single labeler over a single dataset and now utilize multiple labelers and a reviewer over a single dataset, yet have still managed to enhance efficiency.

06 Lessons Learned

Nightfall's experience underscores the importance of flexibility and efficiency in data labeling processes. By adopting a platform that offered customizable features like Reviewer Mode, QA tooling and reports, they significantly enhanced their data quality and operational productivity. This case study reveals that leveraging advanced tooling can lead to substantial efficiency gains, even when scaling up the team involved in the labeling process.

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