

Machine Learning system for integrated loans

Standardized monitoring and deployment of Machine Learning models in production.

Industry: **Banking**

*Image for illustrative purposes only

Summary

Modernization and improvement of Machine Learning infrastructure for a Fintech client, addressing the lack of visibility into production predictions.

Key challenges

- Insufficient visibility and monitoring of production predictions from machine learning models.
- No infrastructure for A/B testing of models and parameters in production.
- The model deployment process was not standardized or well documented.

Approach

Novacomp's team integrated with the company's native team, taking part in their daily stand-ups, sprint planning, and syncs, collaborating with other Novacomp teams such as Data Engineering and Analytics. As a Staff Augmentation service, a state-of-the-art machine learning architecture was designed and implemented using FastAPI, Split.io, Weights & Biases, Docker, and Apache Pulsar.

Results

- Development of an event-based monitoring system for machine learning predictions using Apache Pulsar.
- Documentation of the deployment process, letting the data science team implement changes on their own.
- Improved latency of machine learning services through code refactoring and asynchronous calls.

Technologies used

Python	Catboost	Split.io
Weights & Biases	Apache Pulsar	
XGBoost	Scikit-learn	Flask
FastAPI	Docker	

Conclusion

Implementation of a microservices-based machine learning architecture that enabled parallel model deployments for A/B testing.