

Case Study

AltexSoft Created a Machine Learning Model Behind a Car Insurance Quote Predictor

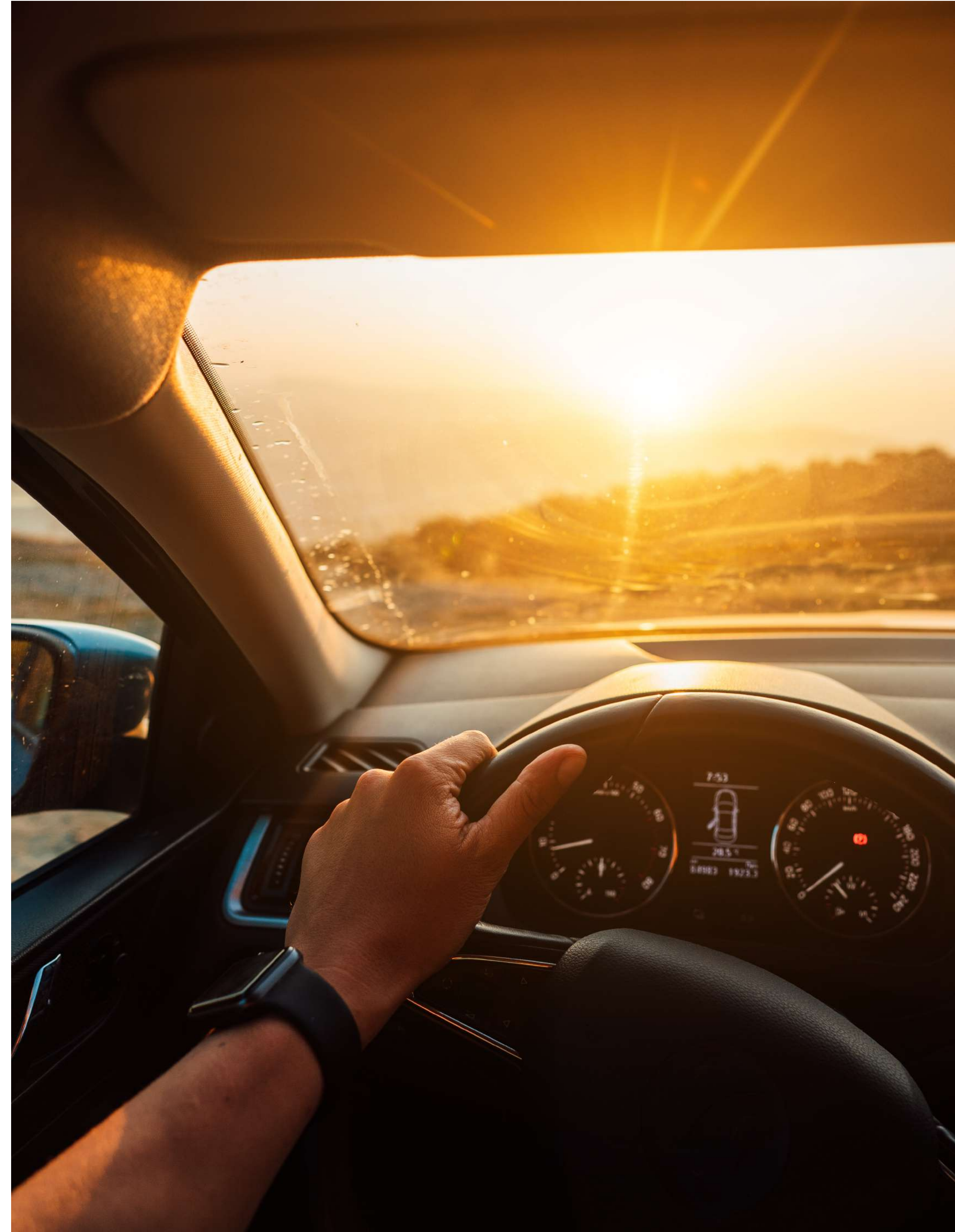
Python, Pandas and scikit-learn libraries, Flask, Docker, LightGBM algorithm



Background

Our client runs a price comparison platform where, among other things, visitors can obtain rates from more than a hundred auto insurance providers. But primarily, users have to complete the online form with around 50 fields, which takes quite a lot of time and discourages many potential customers.

To engage more car owners at the start and increase the conversion rate, the client developed a quote predicting widget with a reduced form to be placed on third-party websites. Based on just a few parameters, the tool forecasts an average cost for insurance and, if the users are satisfied with it, redirects them to the main platform. AltexSoft's data science team created a machine learning model behind the quote predictor.



Challenges

The key challenges AltexSoft faced within the model development process were:

1.

Prepare a relevant subset of quote estimates

2.

Select the best algorithm to forecast average quotes

3.

Integrate the model with the widget

4.

Keep the model updated on a regular basis

Value delivered

1. Selecting a representative subset of data from over one million entries

The initial dataset provided by the client included over one million entries by users, collected via the main platform for several years. However, it contained a lot of excessive and irrelevant details we had to get rid of. Besides that, after doing a lot of experimentation, we found out that several years' data leads to less accurate results than most recent information.

That's because insurance prices are volatile and change over time. So, the final dataset included fresh observations collected over the previous year.

At the end of the day, only about 20 percent of initial data were selected to train a model.

3. Creating a web service to integrate the algorithm

One of the final steps was to integrate the existing algorithm with the client's widget. For this purpose, our team developed a web service, based on the Flask microframework, and deployed it using Docker. It allows the widget to feed new data to the prediction model and retrieve forecasts to be displayed to users.

2. Choosing the fastest algorithm to predict quotes

Random Forest and LightGBM are two ensemble machine learning algorithms our data scientists considered for the average quote prediction. Eventually, LightGBM won owing to its capability of processing large amounts of data in a relatively short period of time. The model was built and trained in Python, using Pandas and scikit-learn libraries.

4. Running automated model updates once a month

The web service also includes a Python script that enables the model to learn from fresh data and automatically update once a month. This continuous adaptation to change ensures relevant quote predictions over a long period of time. True-to-life results elevate the chances of making users visit the main platform, complete the full form, and, eventually, strike a deal with the insurance company of their choice. This, in turn, helps the main platform attract new insurance partners and expand its pool of options for car owners.

Approach and technical info

The project's scope was **2 man-months**. It was completed by one **Data Scientist**.

The technology stack of the project included **Python**, **Pandas** and **scikit-learn** libraries, **Flask** microservices, and **Docker** (for API building and integration). The machine learning approach included the **LightGBM** algorithm.

Cooperation between the client and AltexSoft is ongoing.

AltexSoft US Sales HQ

654 Matsonia Dr,
Foster City, CA 94404
+1 (877) 777-90-97

AltexSoft Global HQ

11/13 Hromadianska Str.,
Kharkiv, Ukraine 61057
+38 (057) 714-1537
solutions@altexsoft.com



altexsoft