CASE STUDY

AltexSoft and Bicycle Blue Book: Automating a Product Catalog and Forecasting a Price Range

Python, LightGBM, Pandas, and scikit-learn

Background

Bicycle Blue Book (BBB), a leading US secondary bike market expert, helps the cycling community quickly and easily upgrade their ride. In addition to offline services, our client launched an online platform where you can find out what's your bike worth, buy and sell used bikes, or trade them in for new ones.

Business Challenges

BBB constantly upgrades the technologies behind its platform to make services even more convenient and reliable. They engaged the AltexSoft data science team to automate their catalog and incorporate machine learning into the price valuation process. These efforts aimed at solving the following business challenges:
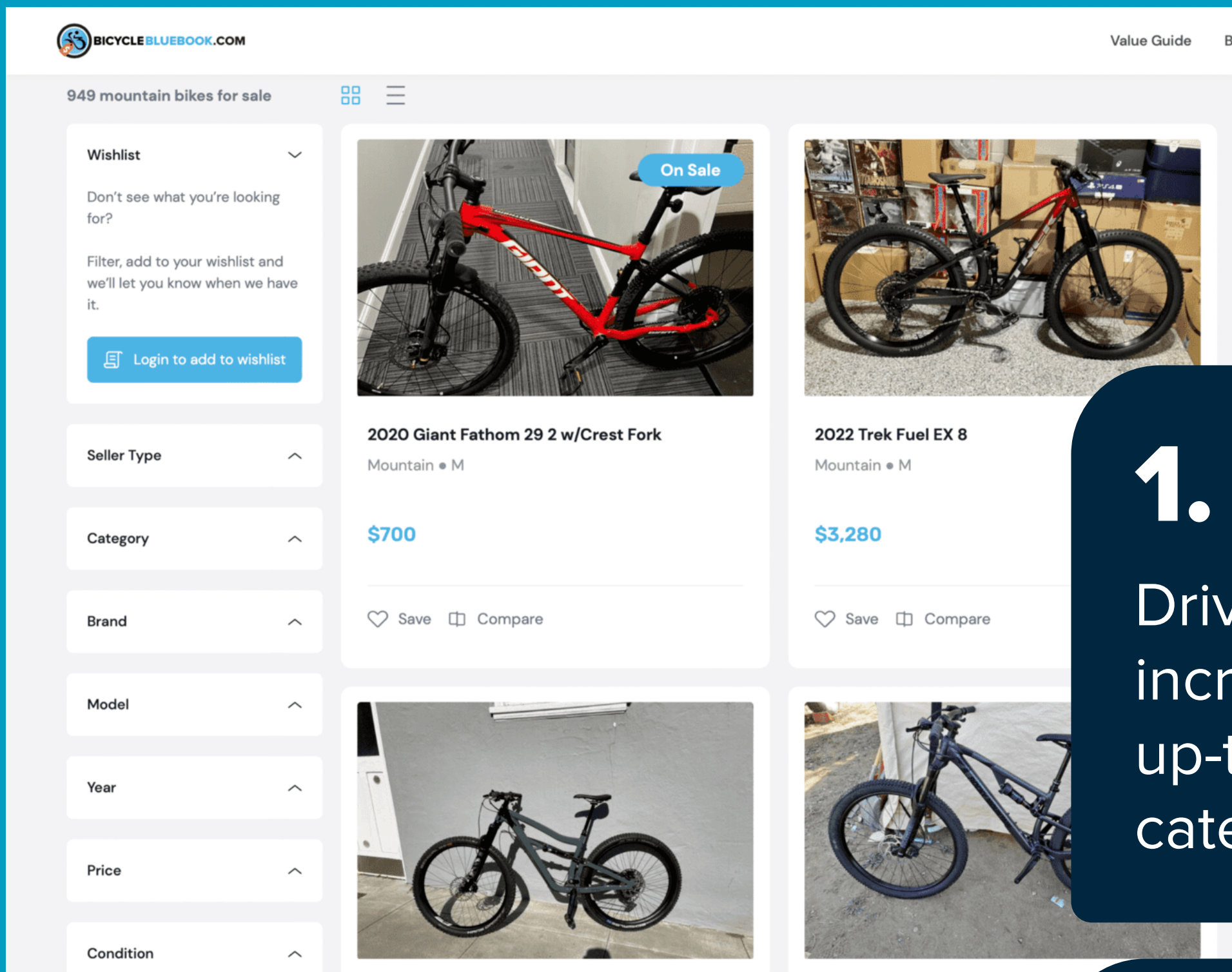
1.

Drive better user experience and increase sales with effective, up-to-date product categorization

2.

Get faster catalog updates and reduce manual work

3.

Enhance the reliability of price prediction for used bicycles

Value delivered

1 **Migrating data to a two-level catalog for flawless product navigation**

BBB maintained a single-level catalog of bikes manufactured over a few decades. As the market changed and new bicycle subtypes appeared, product categories got too broad, each containing thousands of models. Eventually, it became difficult to find a particular bike. So the company decided on modernizing the existing taxonomy. BBB developed a two-level catalog hierarchy and defined attributes important for categorization. We took over the data science side of the project: extracted attributes from the product descriptions using natural language processing (NLP) techniques and cleaned and normalized data. Also, we automated a migration process from the old catalog to the new one.

2 **Automating the import of new products to the catalogand**

The company supplements its inventory with items from various 3rd-party sources — including catalogs by brands. To save time and effort, BBB wanted new products to be added automatically. Since names, categories, and properties in different systems don't coincide, we wrote, tested, and adjusted mapping rules between external and internal catalogs. Finally, we built an NLP-based pipeline that finds meaningful words (brand, model name, etc.) in the bicycle title and other text elements, maps them to unify the content, imports new items to the BBB catalog, and places them under corresponding categories and subcategories.

3   Developing an ML system to predict price depreciation

A core feature of BBB is its online value guide which recommends a price range for second-hand bikes. For more precision, BBB planned to apply machine learning. We chose a LightGBM algorithm known for its high accuracy and nonlinearity. To train it, our team pre-processed data from two sources — BBB and eBay sales. The algorithm takes as input various features including the manufacturer's suggested retail price (MSRP), brand, year of production, type (category), condition, and others. It predicts price depreciation (how much value a bike is likely to lose over a certain period of time) and calculates a price range from the forecast.

Approach And Technical Info

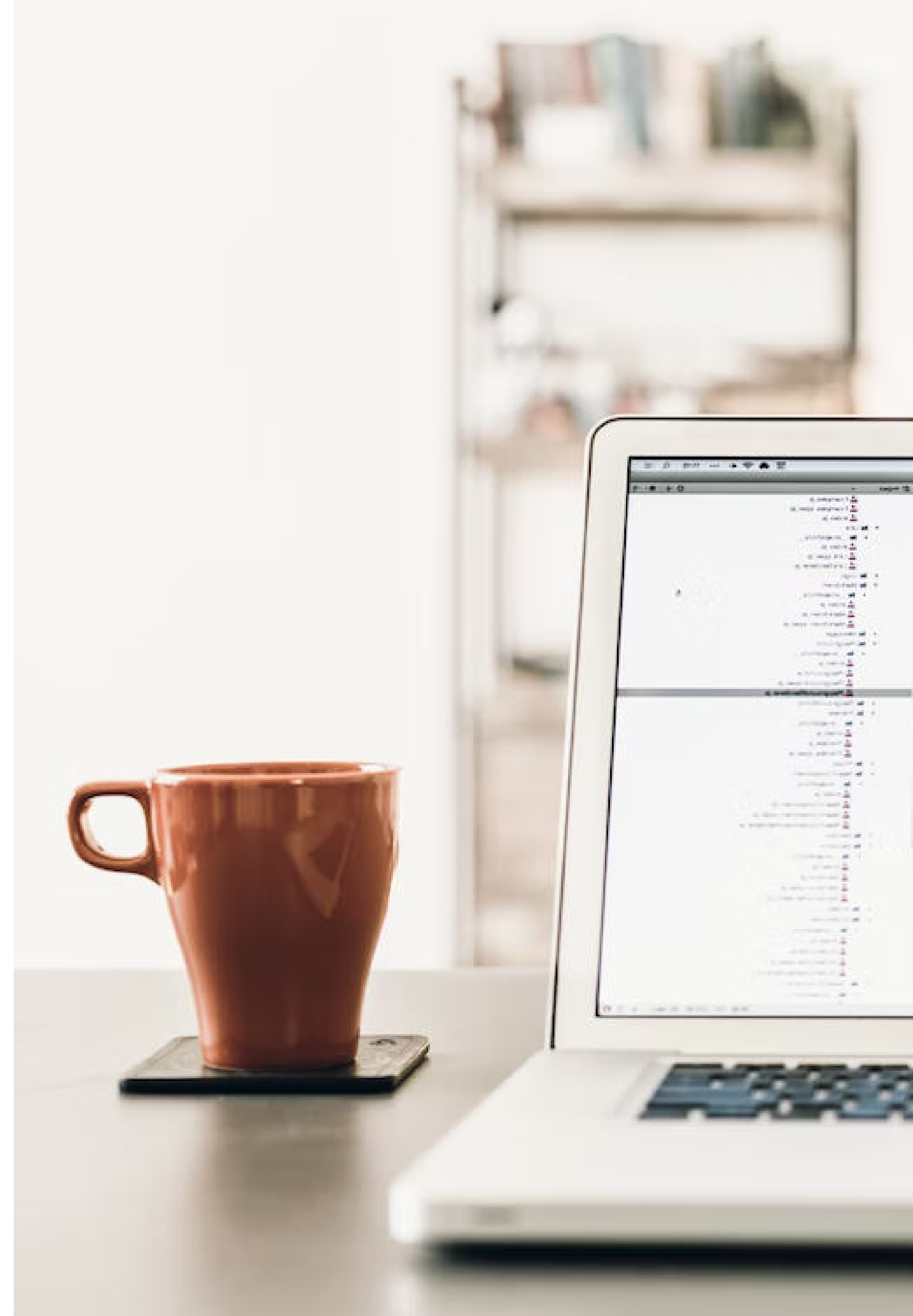A data scientist and a lead data scientist worked on the project for around a year.

The technology stack included:

Python   LightGBM   Pandas

Scikit-learn

Get In Touch With Us

ALTEXSOFT US SALES HQ

654 Matsonia Dr,
Foster City, CA 94404
+1 (877) 777-90-97
solutions@altexsoft.com

ALTEXSOFT GLOBAL HQ

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

altexsoft