OSCORE {ML}

- People
- Tools
- Processes
- Solutions

in
Machine Learning

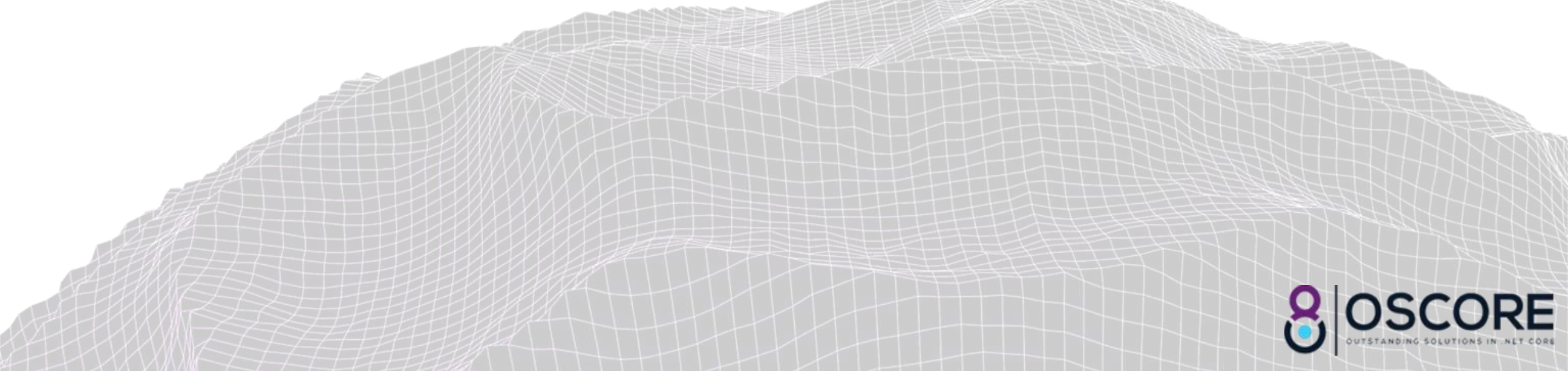
People

Our team consists of…

- Project Managers
- Data Engineers
- Data Scientists
- DevOps Engineers

Expertise:

- PhD in Mathematics
- MSc Artificial Intelligence
- MSc Computer Science
- Microsoft Certified Azure Data Science Associate



OSCORE
OUTSTANDING SOLUTIONS IN .NET CORE

Tools

Cloud Platforms

- Amazon Web Services
- Microsoft Azure
- Google Cloud

Languages

- Python
- C#

Frameworks/Tools

- Pandas
- NumPy
- PyTorch
- Scikit-Learn
- TensorFlow
- SciPy
- Jupyter
- …

OSCORE
OUTSTANDING SOLUTIONS IN .NET CORE

Process

Data Loading and Processing

Data Stores

Synthetic Data Generation

3rd Party APIs

Data Loading

SQL

Azure Data Factory

AWS S3

Google Cloud

Python Scripts

Data Cleansing and Transformations

Encoding, Imputing …

Feature Engineering and Model Training

ML Algorithms – XGBoost, LightGBM, TabNet, Deep/Recurrent Neural Networks…

PyTorch

TensorFlow

Scikit-Learn

Create and Train Machine Learning Models

Feature Selection and Engineering

Feature Selectors

Exploratory Data Analysis

Correlation Matrices

Model Evaluation and Production

RMSE

Accuracy

Precision

Model Execute / Evaluate

Explainability (e.g. SHAP)

Docker

Custom ML Code

ML Model in Operation

Azure ML

Amazon SageMaker

CI/CD

GitHub Actions

ML Training in Operation

Azure Pipelines

Project Management / Client Liaison

OSCORE

OUTSTANDING SOLUTIONS IN .NET CORE

Does This Fit Your Requirements?

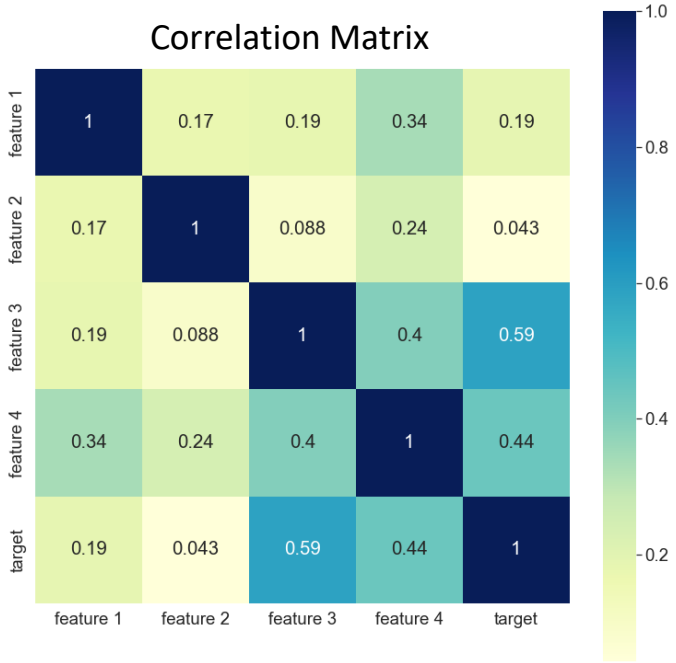
Example Solutions

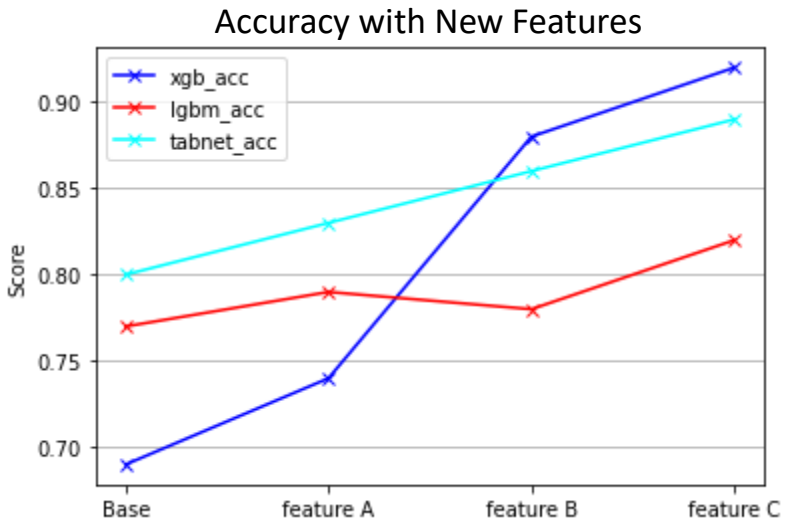
- Solution 1 Retail Unit Closures – *Classification*
- Solution 2 Research Paper Helper – *NLP*
- Solution 3 Stock Market Prediction – *Recurrent Neural Network/Deep Neural Network*

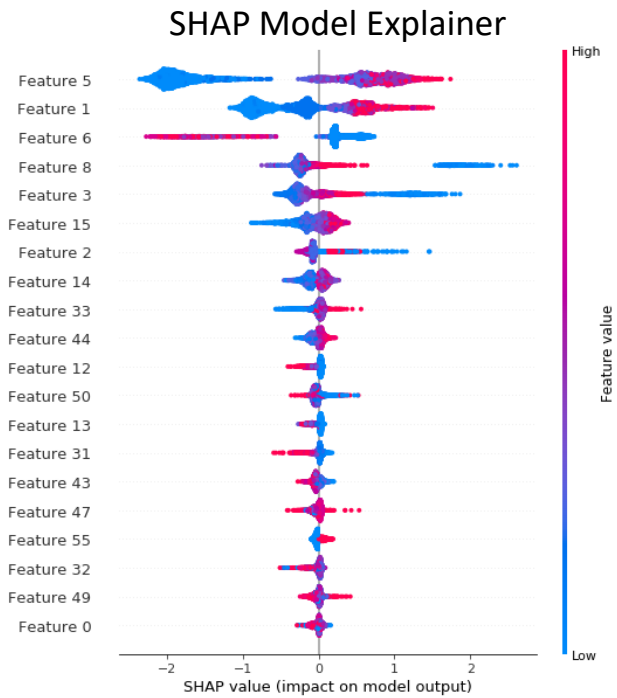
end

OSCORE
OUTSTANDING SOLUTIONS IN .NET CORE

Sample Artifacts

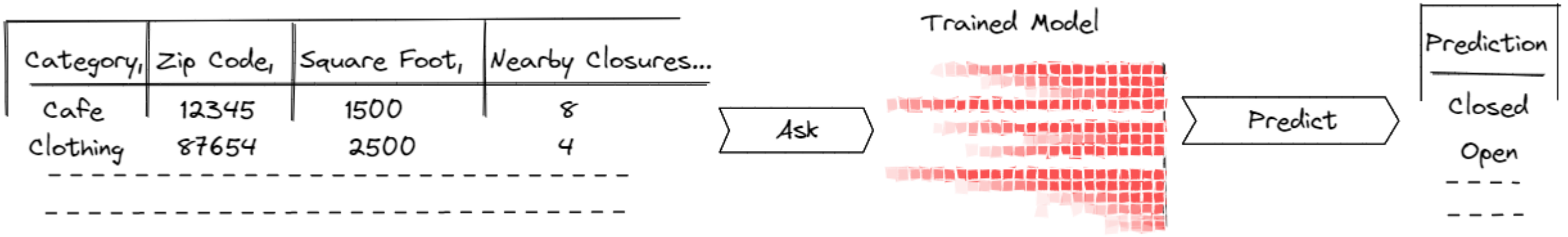






Retail Unit Closures

- Predict whether a retail unit will close or stay open
- Classification problem
- The data included demographic features, geographic features, store categories and more



From client's SQL database

Insert into client's SQL database

Retail Unit Closures

Data Loading and Processing

Data Stores

Data Loading

SQL

Azure Data Factory

AWS S3

Python Scripts

Data Cleansing and Transformations

Encoding

Imputing

Feature Engineering and Model Training

ML Algorithms – XGBoost, LightGBM, TabNet

PyTorch

Scikit-Learn

Create and Train Machine Learning Models

Feature Selection and Engineering

Feature Selectors

Exploratory Data Analysis

Model Evaluation and Production

Accuracy

Precision

Explainability (e.g. SHAP)

Model Execute / Evaluate

ML Model in Operation

Azure ML

CI/CD

ML Training in Operation

Azure Pipelines

solutions

Project Management / Client Liaison

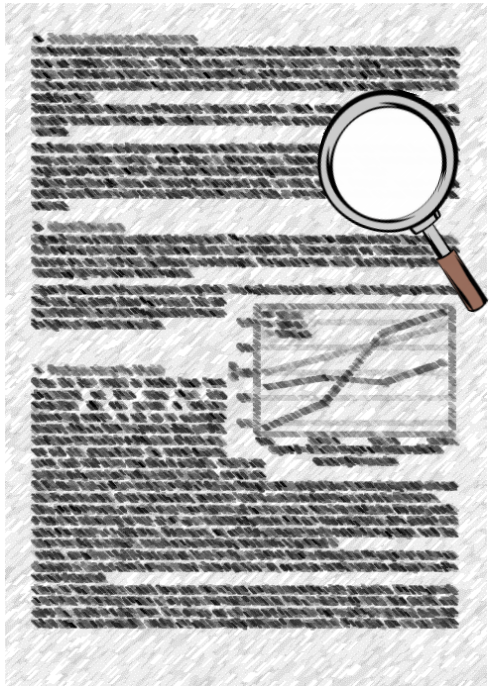
OSCORE

OUTSTANDING SOLUTIONS IN .NET CORE

Project Plan

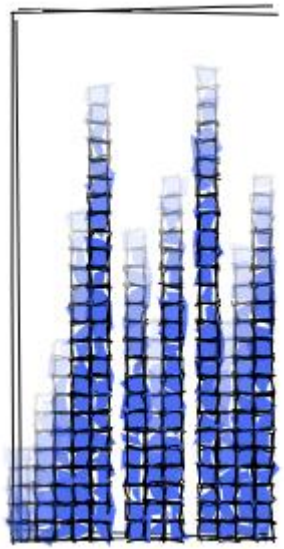
Task	Detailed Steps	Estimated Duration (1 person per day)
1. Data preparation	a. Write SQL queries b. Read from SQL c. Transform into clear tabular form d. Encode and impute values	2-4 weeks
2. Feature Selection and Engineering	a. Determine which features are most important b. Add any new features (discuss with client)	2-3 weeks
3. Train and Evaluate Machine Learning Models	a. Prepare model training steps b. Train multiple models c. Compare predictive performance of models	2-3 weeks
4. Productionize Model and Processing Steps	a. Data processing steps loaded into Azure ML pipeline b. Use best trained model within Azure ML c. Raw data from SQL -> Azure Data Factory -> Azure ML -> predictions back into SQL	1-2 weeks
5. Productionize Training	a. Load training steps into Azure ML b. Model retrains each week	1 week
	Total:	8-13 weeks

Research Paper Helper (NLP)



Research papers (Google Scholar)

Scrape



Python Bot

Returns Text Information

Metric	Page no.	Text
kg/m2 CO2	4	8 kg/m2 of CO2 per day
kPa	9	101 kPa at sea-level
kgCO2	15	6.0 kg CO2 per serving

Text returned in tabular form

Research Paper Helper (NLP)

Web Scraping

Google Scholar

Web search engine

Selenium

BeautifulSoup

Web Scraper

Fetching pdf links and other information

XML

Fetching text and text processing

NLTK

Spacy

Regular Expression

Text pre-processing

Reading text from pdf

PdfPlumber

Tabula-py

Natural Language Processing

Gensim

Creating Word Embeddings

Generating Term – Rank Information

Labelled Training Samples

Named Entity Recognition
Google → Organisation
John → Person

Spacy

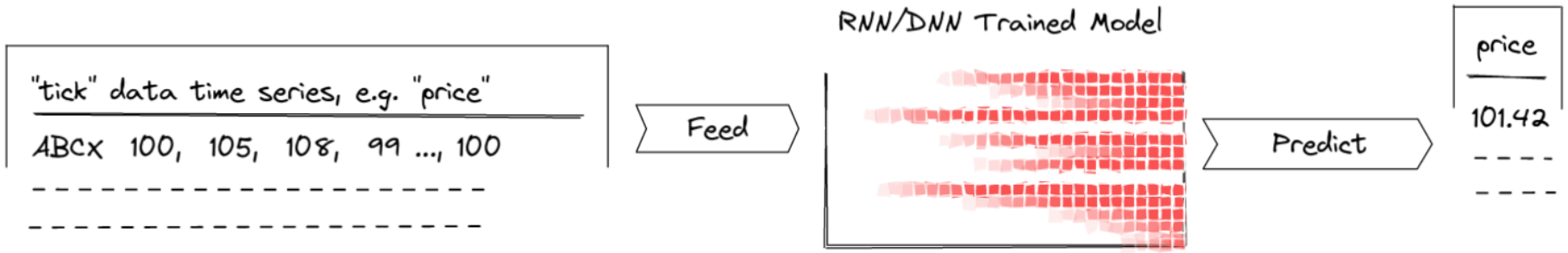
solutions

Project Management / Client Liaison

OSCORE
OUTSTANDING SOLUTIONS IN .NET CORE

Stock Market (Prediction)

- Predict financial markets
- Regression problem on time series
- Using recurrent/deep neural networks



solutions

OSCORE
OUTSTANDING SOLUTIONS IN .NET CORE

Stock Market (Prediction)

Data Loading and Processing

Feature Engineering and Model Training

Model Evaluation and Production

Data Loading

ML Algorithms – Linear Models, Recurrent Neural Nets, Deep Neural Nets

PyTorch

Artificial missing values

$$W_{t+u} - W_t \sim \mathcal{N}(0, u).$$
$$X_t = \mu t + \sigma W_t$$

Create and Train Machine Learning Models

Model Execute / Evaluate

Synthetic Data Generation

Client Supervised ML Research

ML Model in Operation

Standard models of financial market dynamics

Mathematics & Statistics

solutions

Project Management / Client Liaison

OSCORE

OUTSTANDING SOLUTIONS IN .NET CORE

We are Looking for the Next Challenge

Oscore ML

Introduction

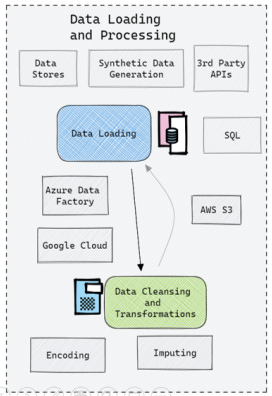
At Oscore ML, we have a dedicated team of machine learning engineers and developers that can understand our clients' requirements and produce robust solutions to AI/ML related problems. We have a wide range of skills and knowledge in software development (especially Python and C#), machine learning, and data science. We strive to provide the best possible services and solutions for our clients.

Approach to Solving a Machine Learning Problem

We aim to deliver the best possible results in the least amount of time. In order to do so, we have a structured process for carrying out a machine learning task.

Initial Communication

Client communication is one of our top priorities, and so we make sure all stakeholders are in the loop every step of the way.
An initial discussion establishes the project needs and gives us an opportunity to explore any available data. This also allows us to consider the technologies that can be used, and we can discuss possible routes to get to the intended solution.
We then produce a project plan consisting of well-defined targets, steps to achieve those targets, and an estimated project duration. After discussing the plan and getting approval, we begin the main phase of the project.



Data Loading

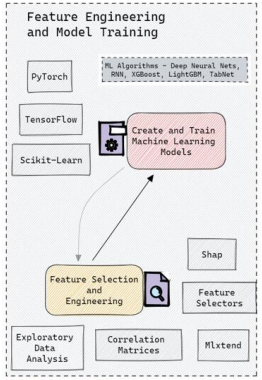
Our machine learning team is experienced in cloud technologies such as Microsoft Azure and Amazon Web Services as well as using SQL. The data used for our ML projects is often loaded from cloud storage, where it is easily accessible and reusable. We store data in the most efficient way to optimize data storage and data loading.

Data Cleansing and Transformations

We also undertake data processing, cleansing, and structuring, to transform the data in preparation for training. This allows us to determine any inconsistencies in the data, such as features with missing or incorrect data, for which we can provide a solution.
During this step we can merge any separate datasets into one, ready for training. These processing steps prepare the data in a repeatable way, so that we can prepare any new data in the same fashion.

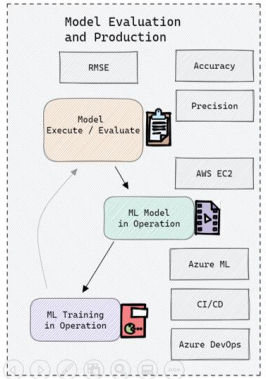
Feature Selection and Engineering

Once the data has been processed, we can look at feature selection and feature engineering. For this stage we use several tools and visualizations to analyse the data such as correlation matrices. These help us to determine any features that are highly correlated with the target variable and any features that have low correlation, and are therefore not likely to affect model output. Feature selectors are another tool used to determine which features are likely to be most important to the model.
This is also a discussion that will take place with our client, where we can look at possible new features that can be added from the current data, or any new data that could be valuable.



Create and Train Machine Learning Models

After we have selected the features to be used, we begin with model training. For building ML models, we use frameworks such as PyTorch, TensorFlow and Scikit-Learn. We first experiment by training using a selection of algorithms that we think are best suited for the problem, this allows us to view the project from a few different angles, and then determine which is the best route to take. We also optimize and fine-tune the hyperparameters of each model when required to improve the overall performance.



Model Evaluation

Once we have trained the models, we begin evaluating model performance to judge which is best suited for the data. For this, we consider several metrics for comparison such as accuracy and precision in the case of classification tasks, or RMSE and MAE for regression tasks. We also explain the machine learning model by plotting feature importance, this tells us which features had the most and least impact on model output, which we compare against the results from our feature selection process. If the results are not sufficient, we may look to collect more data, add new features or retrain the models. At this point we would repeat our feature selection and model training steps.

Productionizing the Model

Once the best model or models have been determined, we then move them into production, where we can provide our client with a real-time inferencing service, or set up an automated process, that can load and process data, predict, and store the results in a data store. If necessary, we will also productionize the model training steps, such that when new data is available, the model is automatically retrained and updated. If suitable, we can use CI/CD tools such as Azure DevOps or GitHub Actions throughout the project to automate deployment into our clients' environment.
Once in production, we are on hand to monitor and optimize the service if required.

Contact us at

ml@oscore.io

for more info

solutions

OSCORE
OUTSTANDING SOLUTIONS IN .NET CORE