accenture

INTELLIGENT LEARNING SERIES: DATA SCIENCE

VIDEO TRANSCRIPT

Hi everyone! My name is Laurence, from Insights and Intelligence here in Accenture Operations.
A data scientist is someone who is responsible for modelling, whether that be for predictive purposes or for a deeper understanding of the drivers affecting the issues or improvements in a given business objective.
To give further details on what we do, I will walk you through our framework of a Data Science project.
I'll mention the core skills of the team, the common tools and techniques that we use, and on the last part, I'll show you some examples of actual projects that we have implemented and their corresponding benefits to Operations.
Now, let's start with the framework.
Before a data science project kicks off, we get into discussions with Operations to know more about their business, to understand their key objectives, and help them identify their pain points.
From there, we will create a project plan which includes the details of the data science solution and present its potential value for them.
Once approved, we can officially start with the project.
It begins with data gathering and data exploration. This also comes with a series of meetings with Operations to clarify questions and align with our initial observations about the datasets provided.
This is quickly followed by the data cleaning and formatting in preparation for modelling.
In this stage, we utilize R or Python programming to conduct possible outlier removal, correction of entries, oversampling or under sampling, and structuring the data into a specific format.
Now, the training, testing, and validation sets will be prepared.
We will then proceed in running a series of models from the set of techniques applicable for the solution. When completed, we will assess the model performance based on evaluation metrics such as Accuracy and F1 score for classification-based models.
Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) for regression-based models.
Afterwards, these results are presented to Operations for both alignment and validation.
When everything is completed, deployment of the model will be facilitated and is subjected to monitoring in the next couple of months.
If no issues are encountered, we will package this model into an executable app so that Operations will be capable of doing the refresh moving forward. That will actually cover the project's end to end framework.
In the different stages of the framework mentioned, two core skills are essential for the team.
First is statistics. Having a good background on Statistics provides us the following:
- It empowers the team to conduct complex data preparation and exploration
- It provides better understanding on what modeling techniques to use
- And it helps us to effectively evaluate the best model to deploy

The second core skill is R and Python Programming.

Copyright © 2022 Accenture
All rights reserved.

Accenture, its logo, and High Performance Delivered are trademarks of Accenture.

With this, the team can do the following:
- First, automate majority of the stages in the Data Science Framework. Whether it be in the data exploration stage or in the deployment stage where an executable app is developed.
- Also, it enables us to access the readily available Data Science packages in R and Python.

Before I move on, I'll quickly give you examples of the datasets that we are working on.
They can be classified into two: Structured Data and Unstructured Data.
Structured data is data in standardized format which consists of values with defined data types. Examples of this are database extractions, and Microsoft Excel data.
Unstructured data is unorganized data that usually deals with sentiments, images, videos, and emails.
Here in our team, we usually work with structured data but on some projects, unstructured data are also present.
Now, I'll go through the tools and techniques that the team usually uses.
For the techniques, we implement a wide range of methods for each type of solution.
For regression-based models, we can opt to use Linear Regression, Support Vector Regression, Random Forest, and Gradient Boosting.
While for classification, we usually use Logistic Regression, SVM, also both Random Forest and Gradient Boosting.
For Clustering, we utilize K-means and Hierarchical Clustering while for time series, we use ARIMA and its variations: SARIMA and SARIMAX, VAR or Vector Autoregression, and Facebook Prophet.
Other techniques that we do are related to text analytics, dimensionality reduction, and optimization.
The main tools that we use are R and Python but some of us have also handled other tools like SAS and Stata.
Finally, I'll be sharing with you two examples of actual data science projects that we have worked on.

The first one is for a network company in Europe concerning the Order Management process.
The client is checking for ways to improve order fulfillment on internet installations. With this objective, the team developed a classification-based model to identify orders that are likely to take longer time to fulfill. Upon achieving an acceptable accuracy and F1 score,
we integrated the output in their order prioritization. This solution resulted to a significant increase in volume of installations and high value of accelerated revenue.
The main tools that we used for this project is Python and Power BI.
For the second project, an Oil & Gas company in US wanted a better foresight of the invoices received on a monthly basis. Given this objective, we developed an ensemble of time series models to predict the incoming volume using the monthly historical data and some industry indicators like oil price.
This project improved the accuracy of their forecast and has provided better capacity planning which in turn reduced their monthly overtime hours. For this project we utilized R and Tableau.
This concludes my discussion for Data Science.

Copyright © 2022 Accenture
All rights reserved.

Accenture, its logo, and High Performance Delivered are trademarks of Accenture.