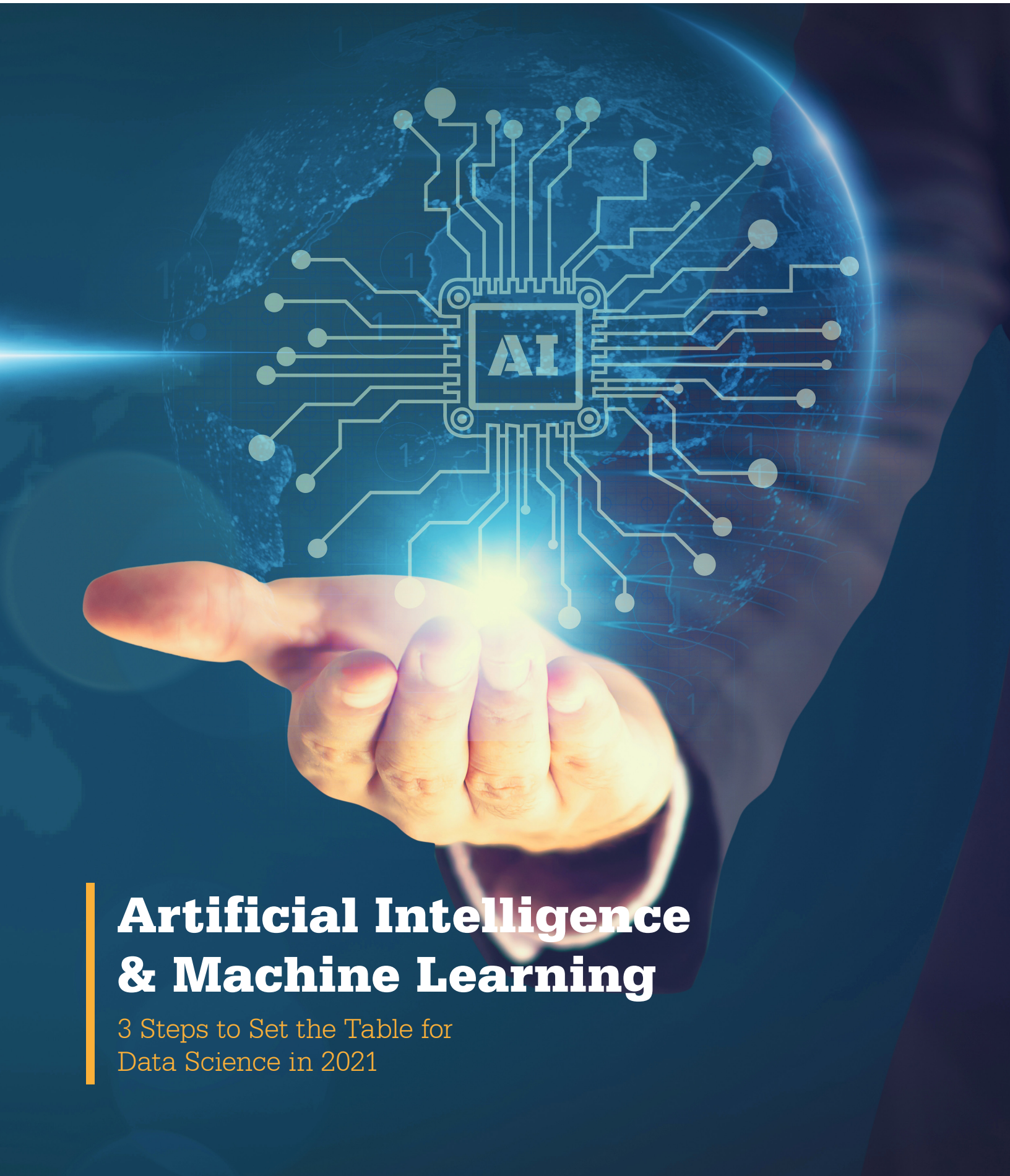


Artificial Intelligence & Machine Learning

3 Steps to Set the Table for Data Science in 2021

Artificial Intelligence & Machine Learning:

3 Steps to Set the Table for Data Science in 2021

Unprecedented events in 2020 pushed organizations into a new normal where predictability, trends of the past, and forecasts went out the window. As organizations navigate 2021, data science, **artificial intelligence** (AI), and **machine learning** (ML) can help you make sense of your data to save time, reduce costs, enable users, and grow profits.

What are AI and ML?

While they tend to go hand-in-hand, AI utilizes technology to do things that typically require human scrutiny, while ML is the application of statistical learning techniques to automatically learn patterns in data. The patterns are used to develop algorithm-based models that make more accurate predictions about the world. Both AI and ML utilize data science to accomplish these outcomes.

Large sets of data that you likely already have, inexpensive storage options, and cloud processing capabilities, along with ML, gives organizations new insights to make data-informed decisions. To guide you through the process of creating, implementing, and replicating a data science project, there are three steps necessary to producing valid and applicable results.

FIND YOUR PURPOSE

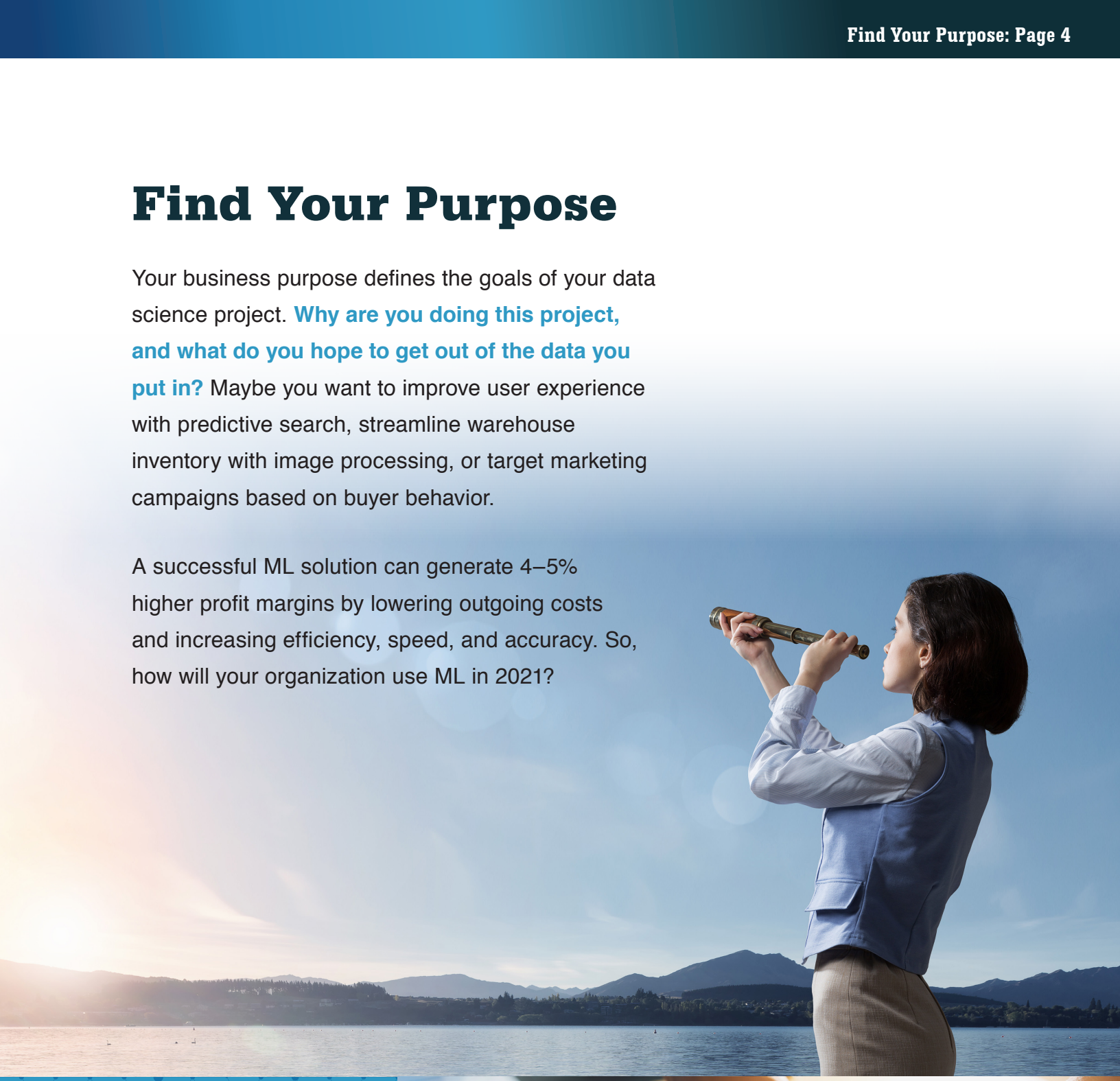
WRANGLE YOUR DATA

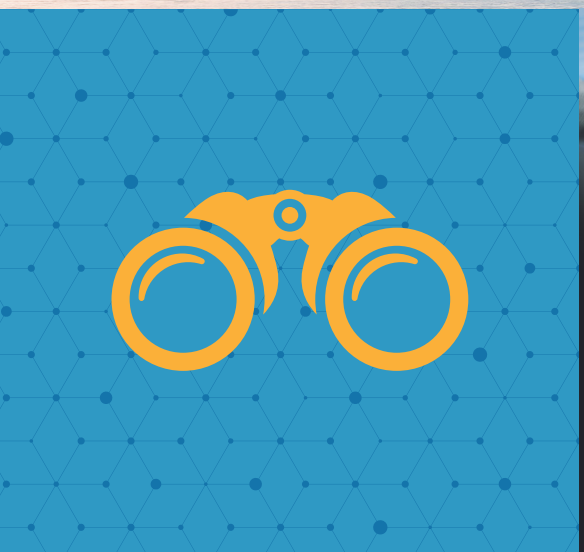
EXPERIMENT & PUSH INTO PRODUCTION

Find Your Purpose

Your business purpose defines the goals of your data science project. **Why are you doing this project, and what do you hope to get out of the data you put in?** Maybe you want to improve user experience with predictive search, streamline warehouse inventory with image processing, or target marketing campaigns based on buyer behavior.

A successful ML solution can generate 4–5% higher profit margins by lowering outgoing costs and increasing efficiency, speed, and accuracy. So, how will your organization use ML in 2021?





Wrangle Your Data

The ML Model

The ML model acts like a program that's trained on sample data using an algorithm to learn patterns. Based on the information learned, the model is applied to inputs and predicts an outcome.

First, the computer gets input data—for example, what you know about customers (i.e. age, demographics, spend, etc.). Next, the computer is fed the output you desire—let's say, which customers churn. Then you build a model by training programmed algorithms to analyze input data and predict an outcome.

The model recognizes what data the output has in common. In this example, it's patterns of customers who churn. The result is common data points observed in customers before the output, churn, is reached.

Wrangle Your Data

Data Preparation

It can be tempting to use raw data, but if it's not clean, governed, or appropriate for use, you won't reach your purpose. Invest in a data cloud, data warehouse or data lake for company-wide visibility. Data should be catalogued, accessible, and searchable in central repositories with data quality and governance applied at all levels. Once the data is complete, correct, and available, your data scientist can start applying ML algorithms.

Wrangle Your Data

Algorithms

Algorithms are available in all the major service provider platforms and many Python and R libraries. General use cases include:

- Classification using anomaly detection, marketing segmentation, and recommendation engines.
- Natural Language Progression (NLP) using autocomplete, sentiment, and understanding (i.e. chatbots).
- Timeseries using forecasting.

POPULAR ALGORITHMS

CNN
convolutional neuro network

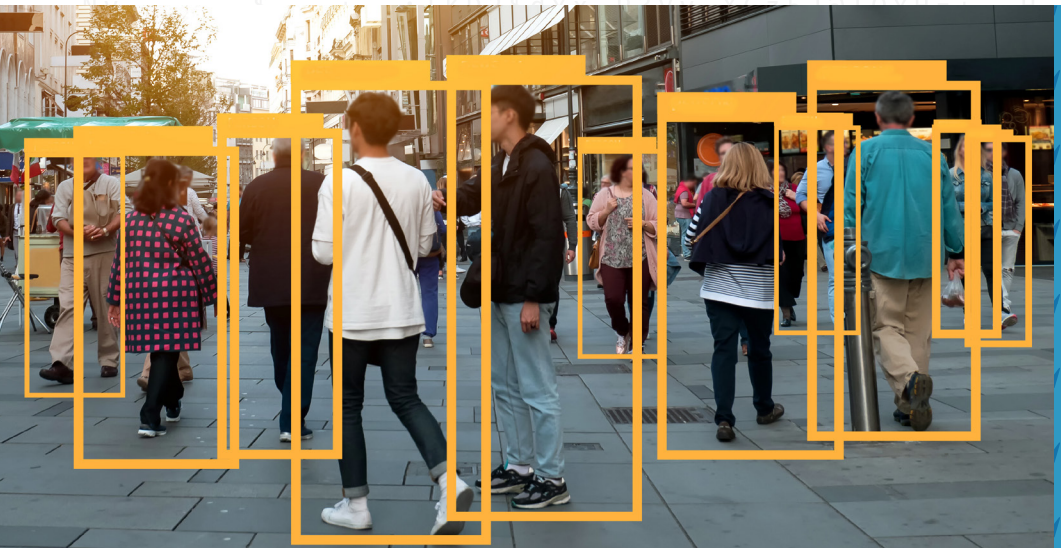
K Means Clustering

PCA
principal component analysis

Support Vector Machine

Decision Trees

Logistic Regression



Wrangle Your Data

Model Quality

Model quality depends on bias and variance. Bias quantifies the algorithm's limited flexibility to learn patterns. Variance quantifies the algorithm's sensitivity to specific sets of training.

COMMON MODELS

1

OVER-FITTING

LOW BIAS + HIGH VARIANCE

This model is tightly fitted to training data and won't generalize new data.

2

UNDER-FITTING

HIGH BIAS + LOW VARIANCE

This model is new and hasn't reached a point of accuracy. Get to over-fitting first, then back up and reiterate.

3

LIMITING/PREVENTING UNDER/OVER-FITTING

There are too many features, or data points, in the model. Reduce them or create new features from existing features.

Experiment & Push into Production

Data science requires you to see, touch, and feel the results to know if it's working. Luckily with the cloud, the cost of experimentation is low.

Once you're in production, promote your project around the organization and garner attention from executive leadership. Unfortunately, 70% of data projects fail because they don't have an executive champion, so share what you learn with visual representation, a focus on company-wide impact, and emphasize the need to repeat and the budget necessary to do so.

Conclusion

If 2020 has taught us anything, it's that you can't predict the future. However, machine learning and data science can help you decipher the past to create a better future. Streamline the process of introducing, completing, and repeating a data science project using best practices.

With 2nd Watch **Data and Analytics Services**, you realize the power of machine learning with the right algorithm selection and model deployment.

Want to dive deeper?

Contact us to see how data science, machine learning and artificial intelligence can impact your organization.