Getting Started with Machine Learning

Machine learning is now easily accessible for companies of all sizes due to the recent availability of robust data platforms. And while these platforms make it faster and easier to add ML into your business, a successful implementation still requires carefully planning.

For more information about Machine Learning concepts, see **tallan.com/ML**

We've prepared this step-by-step guide to help make your first ML projects successful.

What is Machine Learning?

Machine Learning has been around for decades, so why it is just now so pervasive in articles discussing trends for advanced analytics? Recent availability of data platforms, powered in many cases by the cloud, make it possible to implement Machine Learning projects on the massive scale they require, but easily and affordably for companies of all sizes.

Machine Learning solutions fall into a couple of classes, but at a high-level, they all consume large quantities of historical data, and pass that information through proven mathematical algorithms to identify patterns in the data that will lead to a reliable prediction. This creation is called the "model", something we have all experienced at least indirectly, probably most familiarly every time we watch the weather forecast.

In addition to the substantial number of data points or features required to make accurate predictions, the variability and noise in that data is what really contributes to the need for massive data volumes, and in turn require the computational power that only machines can reasonably achieve.

Meet the Author

Jared Brown, Vice President, Data & Analytics
jared.brown@tallan.com

With 18 years of experience, and a strong data engineering background, Jared has been instrumental in the design and delivery of transformative data projects across several vertical markets. Under his direction, the Data & Analytics practice at Tallan has renewed their focus on the modern data platform, to enable all business initiatives with data driven results. This includes everything from operational workloads and data warehousing, to real-time embedded analytics and machine learning.

Step 1: Envisioning - Identify Opportunities

It can be tempting to jump right into Machine Learning just for the sake of using the technology. But to appreciate the power and value of ML, it is important to find the right opportunity. You may have a good option in mind already, but if you are unsure, here are some questions to start identifying areas where ML can benefit your business:

Where can you get value from classifying and predicting behaviors?

ML Classification models can help you categorize your data, which in turn can help predict outcomes. For instance, you could classify customers as likely to leave or likely to stay. With this information, you can change how you interact with these customers, and thereby influence their outcome. Other examples of classification are:

- Customer will (or won't) purchase
- Customer will (or won't) return goods
- Customer will (or won't) use a coupon
- Device will (or won't) fail
- Patient is (or isn't) likely to need readmission

Tallan Recommends: Two-Class Decision Trees are a great introduction to classifier methods. They use an intuitive work-flow-like model, so you're not stuck trying to optimize a black box.

What figures that you directly control influence your business?

ML Regression algorithms can be used to predict the optimal values to use when making decisions that have complex impacts on your business system. A common application is to choose the optimal price point for a product. A more advanced version of this can vary the price point based on multiple factors, such as time of day, demographics, and lead source. Other uses for regression models include:

- Staffing levels
- On-hand inventory levels
- Coupon amounts
- Maintenance schedule

Tallan Recommends: Linear Regression is a great starting point for regression methods. Since you can also create these in Excel, you can easily validate your data before moving into more complex predictions.

Do you have outliers in your data?

ML Anomaly Detection models are excellent for identifying irregularities in your data. The most well-known application of this is fraud detection, where financial institutions can identify irregular behavior seemingly instantaneously. Other uses for anomaly detection include:

- Human error
- Manufacturing batch defects
- Changes in customer behavior patterns
- Device tampering

Tallan Recommends: PCA-Based Anomaly Detection is a great catch-all anomaly detection method. It works by picking out the natural variance in your data, and highlighting anything outside normal.

Step 2: Define Value Hypothesis

Every business is different, so it is important to view your ML effort in terms of the value that it will provide to your particular business.

Document current process

Document the current process that you are interested in improving. This can be fairly high level, but should include where resources (time, money, etc.) are spent, and where value (sales, loss prevention, customer satisfaction, etc.) is gained. Identify where you believe there is room for improvement.

Document the expected new process

Show where and how you will improve the process using ML. Remember, this is still a hypothesis, so it's expected to be unsure whether or not you can achieve what you're attempting. This new process should include feedback loops to improve the model. Also identify any data that you feel would be important for this process, even if you don't currently have access to this data.

Prioritize your efforts

You will likely have multiple areas now where you expect you can improve your process. Pick one that you feel is achievable and valuable, and start with that. If you're not sure what is achievable, start with the most valuable, and re-evaluate it after step 3.

STILL NOT SURE?

If you find that you're stuck, or unsure of how to proceed, it may be time to call an expert. Getting your project started in the right direction can be quick and inexpensive, and it will help ensure the rest of your investment is well spent.

For more information contact us at **analytics@tallan.com** or visit **tallan.com/ML**

Step 3: Collect Data

Machine learning is driven by data, so it is ideal to have good data, but more important to have enough of it so that your model can learn. The data doesn't need to be perfect; machine learning is built around the idea of dealing with larger, but less perfect, data in order to deliver a predictive solution.

Consider the following suggestions:

- Partially-summarized data is good; you can give yourself plenty of data points to work with, while still providing some human feedback to train the process. Keep records at the event level, but augment data with:
    - Summary information such as purchases in past year
    - Interval data, such as time since last purchase
    - Combined data, such as account balance
    - Scoring data, using existing measurement techniques
    - Segmentation data, such as industry or profession
    - Demographic information, such as age, gender, location, etc.
- Don't feel pressured to include everything! Often, some information just isn't helpful, or isn't as helpful as a related piece of information. You can save some stress by keeping things trim early.
- Collect additional data from your systems, and 3rd party data sources.

You may want to also consider topics like data governance (controlling who can see certain data) and better ways to store or track your data. These may not be important topics to address right away, but these discovery stages often expose other things you may wish to improve. Some things to consider:

- New regulations require you to be very careful where and how your data is stored
- Take the opportunity to identify data lineage and chains, and consider where used
- PII and PHI concerns may be present at some levels of granularity but not others, especially as disparate data is being combined.

Valuable information is all around you. Every interaction with customers, partners, and employees can yield valuable data. Much of this information is regulated, but can still be used if properly stored and shared. Tallan offers custom development services to assist you in responsibly collecting, storing, integrating, and using this information.

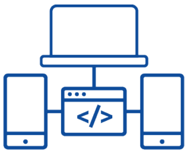
For more information contact us at **analytics@tallan.com** or visit **tallan.com/ML**

Step 4: Test your Hypothesis

Now is where the real fun starts! To test whether you can use ML as predicted, you will create a model and train it with your data. This is easy and fast to do with today's toolsets.

Pick a Platform

First, you'll need a tool to actually implement your experimental machine learning models. There are many out there, but we recommend starting with Azure ML Studio. This platform has a simple user interface, support for models of various complexity, and is free to try. For this section, we will assume you are using Azure ML Studio, but most platforms operate in a similar way. If you are new to Azure ML Studio, we suggest watching the 'Experiment Tutorial' to get started



Create Your Experiment

Now that you've gotten your platform, you'll likely need to explore some of that platform's resources for training and basic use. However, they should all have steps similar to these:

1. Create a new dataset by uploading a local file. You will typically start with a single data set, with all of your sources combined. More advanced techniques can combine data in the experiment.

2. Create a new Experiment. This is just a container for the work you will be doing to train and test your model. ML Studio also has the concept of a Project, which packages the data, experiment, and APIs together.

3. Choose your data. This is where you may have to play around a bit. The data visualization and statistics in ML Studio is useful when working this out. Using more elements is not always better, and can create problems later on if the data is not accurate or complete. Be sure to split off 20% of your data to test with, which can be performed directly in the experiment.

4. Add in an appropriate model. Start with something simple, and try different types if the results aren't what you want. Models are typically classified as:

   a. Classification (identify if something is X, Y, or Z based on provided example)
   b. Regression (identify what value between A and B provides the desired outcome)
   c. Clustering (categorize your data into similar groups)
   d. Anomaly Detection (identify items in your data that don't appear to belong or fit)

5. Train and test your model. When you execute your experiment, it will run the inputs, along with any expected outputs that you provide, into the model so it can learn the right answers. This is where the magic happens! When you score your model using the other 20% of test data, you can evaluate the output and see how well it did.

Refine Your Experiment

Review the output of the model test. ML Studio has a nice visualization page, which tells you how well the test data did when run through your model. Remember that machine learning is not meant to be perfect, but rather just to be close.
If the first attempt didn't give you quite what you'd hoped, there are some simple things you can refactor to get right back on track:

First, try variations on the model type that you've picked. Each model has its own way of learning the information that you are giving it, and some fit better with certain types of data. Your data may have a quirk or distribution that fits better in a different model.

Second, once you have a good model, you can tune it to get a little better. Certain algorithms, for instance, can have settings that limit their complexity. Research how your model can be adjusted, and try making some educated changes.

Finally, if you are still not getting the results you expect, double-check that your data doesn't have serious quality issues in it. If you can't find anything, it may be time to ask an expert to help you get your ML project back on track.

Step 5: Review the Predicted Benefit

Once you have the ML model working as you expected, you should take a step back and ensure that these changes will still provide the business value that you expected. Sometimes the accuracy of the model is not enough to get the business impact that you want, or perhaps the way the data is clustered doesn't add the value you had hoped. Now is the time to find that out.

The most straightforward way to do this is to use a sample set of historical data. Run the data through your ML model to see what the outcome is, and look at individual interactions to see how they will be impacted. You will likely need to consider this in the context of your new business process, which is driven by ML. But at this point you don't need to actually go through the entire process.

If you are satisfied that you are now seeing the results that you had hoped for, you can move on to testing with live data. Otherwise, you should go back to step 4 with your new information and try a different type of model or set of attributes.

Step 6: Test the New Business Process

Before you invest in putting your new business process into production, you want to be sure that the final outcome will produce the value that you expect. In most cases you will need to include live data and actually act on it. The process at this point is usually partially manual, as you haven't fully integrated it into your business process yet.

One easy way to get the data into an actionable format is to use a web API to connect with Excel. This allows you to quickly combine your data from a reporting system with the intelligence from the ML model. An example of this would be to combine customer data with the output of an ML model predicting customer churn. You could then act on a subset of this group to see if your action influences them to stay.

This example shows why it is advantageous to use slightly aggregated and denormalized data for your model when starting off. It would be much more difficult to test with a complex and atomic data set.

Step 7: Integrate into Business

If your test in Step 6 proves the business process works as you expected, you can now take the step of fully integrating into your standard business process. This will be highly dependent on what your process changes are, and will likely include custom software development.

Some good examples of integrating ML into your business process include:

1. Exposing customer categorizations or risk in real time during customer service calls
2. Changing automated marketing processes or target audiences
3. Optimizing inventory levels and ordering amounts
4. Providing personalized offers to customers currently engaged on your site

Each of these scenarios will likely require modifications to your daily use systems. Once you have proven the value out, you can make a well informed decision as to whether that integration is worth the cost. It is important to understand the costs as well as the benefits of the full ML solution.

Once you have integrated the improvement into your business process, or determined that the cost is not worth the potential benefit, you can return to step 2, and review your list of priorities. After a few ML initiatives, be sure to go back to step 1, and rethink your envisioning from a new perspective.

Conclusions

We hope this guide has helped you get started on your journey to improve your business using Machine Learning. The variety of new tools make ML accessible and affordable to every company now, but it can still be overwhelming to get going down the right path. If you get stuck at any point, reach out to an expert and get your project back on track. But most importantly, don't give up, and don't be afraid to fail on a specific path or experiment. Your information holds vast amounts of value for your company, and it's worth putting in the effort to get it.

About Tallan

Tallan has been in the business of helping companies better leverage their data since 1985. Our rapid envisioning team can get you started in days, and our data platform team can guide you in every detail. When ready to deploy, Tallan has expertise in customizing your existing business applications to support the new business process, even if those applications are customized or home grown.

TALLAN

Machine Learning Worksheet

1. Envisioning Session

Opportunities Example: Predict when customers are likely to leave.

2. Define Value Hypothesis

Example: "Win back our valued customers before they even know they plan to leave!"

Current Process Example: Ineffective marketing campaigns targeted exclusively towards new customers

Future Process Example: Focus on customer retention, and specifically customers that are likely to leave, to minimize churn by 10% and increase top-line revenue by at least 20%.

3. Collect Data

Example Sources:

Transactional System (purchase history, geographic/demographic info)

Customer Service Platform (# returns, # calls, resolution satisfaction)

Website (# visits, product reviews, loyalty program)

Social Media (follow, # liked posts, sentiment of comments)

4. Test Your Hypothesis

Platform	
Dataset	
Experiment Name	
Features	
Model Type	
Testing Outcome	

5. Review Predicted Benefit

Future Process with expected benefit based on model results

Example: The tested model suggests it will identity a narrow pool of 15% of customers that are likely to churn, targeting just that group will allow us to include a promotional incentive that should get us to our target 10% retention goal.

6. Test the New Business Process

Test Approach

Acceptance Criteria

7. Integrate into Business

System Integration Plan

8. Next steps

Refine Model Options

Next Objective