WISEITECH: Machine Learning Automation Platform “WiseProphet”

In this issue

Research From Gartner: Data Science and Machine Learning Solutions: Buy, Build or Outsource? 2
WISEITECH: Machine Learning Automation Platform “WiseProphet” 12
Company Information 24
Any new data science or machine learning project presents the decision to build, buy or outsource a solution. This note presents a series of key questions and selection criteria to guide data and analytics leaders toward the best approach as they pursue new analytics and BI strategies.

**Key Findings**

- When getting started with machine learning and data science, many data and analytics leaders are overwhelmed by the sheer number of available solutions, and struggle to understand the differences between them.

- Data and analytics leaders are often unaware that benefits and challenges can vary significantly between the buy, build and outsource solution paths. Every project is different, and many organizations have a hybrid portfolio of built, bought and outsourced solutions.

- Choosing the wrong approach (or mix of approaches) can result in lost opportunities, wasted money, missed deadlines, suboptimal solutions and even catastrophic project failure.

**Recommendations**

For data and analytics leaders responsible for data science projects as part of analytics and BI solutions:

- Pursue the “buy” option when available packaged applications or commoditized machine learning APIs are well-suited to your use case(s), and when cost, accessibility
and time to solution are more important than differentiation or extensive customization. Always explore buy options first.

- Pursue the “build” option if the proposed project would be a critical differentiator for your business, you already have strong or budding data science skills in-house, and/or you require a high level of agility and granularity of control.

- Pursue the “outsource” option if you lack the in-house talent to build your own solution and pretrained models and packaged applications don’t meet your needs, or require prohibitive customization to do so. Also consider the outsource option to temporarily augment data science staff and get new projects off the ground.

- Plan to support a combination of all three options in the long term. The optimal mix will be determined by the relative effectiveness of buy options, the feasibility of build options, and the expense and sustainability of outsource options.

**Strategic Planning Assumptions**

By 2020, 85% of CIOs will be piloting AI programs through a combination of buy, build and outsource efforts.

By 2022, enterprise AI projects with built-in transparency will be 100% more likely to get funding from CIOs.

By 2024, a scarcity of data scientists will no longer hinder the adoption of data science and machine learning in organizations.

**Analysis**

**Introduction**

Data and analytics leaders nearly universally recognize the importance of data science and machine learning solutions as part of their larger AI strategy. Still, most are just embarking on their first project or are looking to expand and connect pockets of data science activity in the organization. Technologies such as predictive analytics, machine learning and deep neural nets (deep learning) sit at or just past the Peak of Inflated Expectations on related Gartner Hype Cycles. While some degree of backlash in the face of overinflated value propositions and underwhelming early results is inevitable, overall an increasing number of organizations are achieving tangible benefits from data science initiatives, and investment in solutions is continuing to grow.

Three basic approaches to obtaining machine learning and data science solutions — buying, building or outsourcing (see Table 1 and Note 1) — will look familiar to IT leaders. However, the variety of options and volume of considerations can be overwhelming. Before diving into the finer details of each option, Gartner offers a simplified decision tree to help data and analytics leaders visualize the core questions that go into beginning a new data science and machine learning project (see Figure 1 on page 4).

The early stages of a project will not always be this straightforward. Data and analytics leaders should consider the subtleties and peculiarities intrinsic to data science and machine learning. To choose the best solution path for acquiring and implementing data science capabilities, data and analytics leaders must understand:
The expected (or required) business value to make the effort worthwhile

The opportunities, realities and constraints of the business problem at hand

Your IT organization’s level of analytics maturity

The availability and scale of skilled staff

The required time to solution

Whether the need is urgent or business-critical

How much agility and control you need

Your organization’s short-term and long-term appetite and budget for data science and machine learning

The availability of specialized and readily available tools — also called packaged applications

The availability of pretrained models and machine learning services — also called machine learning APIs

In the course of deciding whether to buy, build or outsource discrete data science projects, data and analytics leaders should expect to develop a portfolio of all three approaches over time (see Table 1). Buy options should as always be considered first. If sufficiently effective, they offer the lowest total cost and fastest time to value of all three options. However, leaders should plan to build or outsource any projects where they hope to make the application of data science a competitive differentiator. It may be difficult to build projects at first, as internal data science teams and supporting resources are still being stood up. Many organizations hope to reach an end state where they are no longer dependent on outsourcing for data science work and can leverage service providers tactically to augment internal staff when necessary. The optimal mix of the three options will be different for every organization.

Figure 1. Decision Tree for Buy, Build or Outsource
Table 1. Choosing the Right Solution Path

<table>
<thead>
<tr>
<th>Solution Path</th>
<th>Circumstances and Characteristics</th>
</tr>
</thead>
</table>
| Buy packaged analytics applications or pretrained models, even though they may require some adaptation | - Well-defined and prototypical domains, for which vendors offer dedicated packaged analytics applications or pretrained models (available via APIs) that solve a specific problem. This software segment is also called off-the-shelf analytics applications and pretrained models.  
- Access to API libraries/marketplaces and/or cloud-based machine learning services (e.g., Amazon, Google, IBM, Microsoft), as well as AI developer toolkits to integrate ML output with existing or future applications.  
- Existing familiar enterprise applications, such as ERP, CRM and HCM, with embedded or integrated AI technologies, where AI is designed to run unobtrusively in the background.  
- Examples: marketing campaigns, price optimization, social network analysis, condition-based monitoring and fraud detection. |
| Build a solution using data science and machine learning platforms, citizen data science tools and open-source technologies; creating a permanent team for in-house projects | - Skilled staff with a deep understanding of how to build advanced analytics solutions based on data science platforms and citizen data science tools.  
- Tools that range from numerous programming languages (e.g., Python and R) to tools catering to citizen data scientists (e.g., Alteryx, DataRobot, H2O Driverless AI) to various end-to-end platforms.  
- A flexible environment that can handle new technologies, libraries and packages from the open-source community.  
- Prototype environments in which data science professionals can experiment with open-source technologies until the team is ready to package new models for production, or maintain models that require vendor support to operationalize. |

Continued on page 5
Continued from page 4

<table>
<thead>
<tr>
<th>Solution Path</th>
<th>Circumstances and Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outsource building, deploying and/or managing a solution to an analytics service provider or freelancer</td>
<td>A complex or unique data science initiative coupled with a lack of adequate staff to build a solution internally. Gartner defines four major categories of vendors among the myriad service providers claiming that they can build data science solutions:</td>
</tr>
<tr>
<td></td>
<td>■ Global consultancies and system integrators — Multinational organizations that offer a broad range of business consulting and technology integration services, and that incorporate a specialist analytics practice. Examples include Accenture, Capgemini, Cognizant, Deloitte, IBM Global Business Services, Infosys, KPMG, PwC, Tata Consultancy Services (TCS) and Wipro.</td>
</tr>
<tr>
<td></td>
<td>■ Specialist midsize consultancies — These analytics service providers operate in multiple markets, but focus specifically on a range of analytics and ML solutions and services. Examples include Affine Analytics, Fractal Analytics, LatentView Analytics, Mu Sigma and Tessella.</td>
</tr>
<tr>
<td></td>
<td>■ Local companies and analytics solution specialists — Companies that focus on a particular geographic market or on specialized machine learning solutions for a specific industry or business process setting. Representative examples include Blue Yonder, CognitiveScale, Comma Soft, Expert System, H2O.ai and QuantumBlack (see KDnuggets for an unqualified compilation).</td>
</tr>
<tr>
<td></td>
<td>■ Analytics consultancy service brokers/crowdsourcing platforms — Companies such as Kaggle, Experfy and Topcoder act as a middle agent to bring together clients that require specialist input and analytics resources to work on a project-by-project basis. Various individuals and teams list their talents on these marketplaces, and select the kind of data science problems that they want to work on. Interested organizations can search for talent based on past experience or specific skills and technical expertise. Levels of skill range from top-notch (typically from major universities) to beginners enrolled in undergraduate statistics, computer science or machine learning courses.</td>
</tr>
</tbody>
</table>

Source: Gartner (October 2018)

Each of these high-level approaches has strengths and weaknesses. Figures 2 and 3 offer an approximate assessment of how the various categories of solutions score against six evaluation criteria:

■ Ease of use
■ Time to solution
■ Solution quality
■ Cost-effectiveness
■ Learning experience
■ Agility/granularity of control
Figure 2. Evaluating Potential Data Science and Machine Learning Solutions (Buy and Build)

The three colors represent the relative strength of the different solution paths across the six evaluation criteria (green = strong, yellow = fair, red = weak). For example, the ease of use for data science platforms is weak because these platforms typically have a medium to very high skills level requirement. This contrasts with packaged analytics applications, whose ease of use, compared with all other approaches, is typically very strong.

Source: Gartner (October 2018)

Figure 3. Evaluating Potential Data Science and Machine Learning Solutions (Outsource)

The three colors represent the relative strength of the different solution paths across the six evaluation criteria (green = strong, yellow = fair, red = weak).

Source: Gartner (October 2018)
The following eight questions should also guide more detailed internal discussions devoted to deciding which solution path is best-suited to your organization.

1. Is there a packaged application or pretrained model available that solves the analytical problem at hand?

This is the first question you should ask when tackling a new analytics problem, as the “buy” option is typically the fastest and cheapest route to applying machine learning and data science. APIs help developers rapidly access pretrained machine learning models and seamlessly integrate them into applications for consumption by business users. Packaged applications can be implemented quickly and at relatively low cost as point solutions. Leverage your current vendor relationships to evaluate their offerings and explore bundling options. Review “Hype Cycle for Customer Experience Analytics, 2018,” “Hype Cycle for Back-Office Analytic Applications, 2018” and “Comparison of Amazon, Google, IBM and Microsoft AI Cloud Services: 2018 Update” for an initial overview of the technologies in this category. Use the Hype Cycles to learn naming conventions, clarify definitions and descriptions, and look up analysts covering various topics.

2. Is there a packaged application or pretrained model available that meets many of — though not all — your requirements?

Even if there is no packaged application or pretrained model available that is perfectly suited to a particular analytical scenario, the buy approach can often provide a “good enough” solution that delivers rapid ROI. Time to solution is a strength for the buy option, which can minimize opportunity costs. Packaged applications and pretrained models can also serve as a stopgap until other solutions can be developed and implemented. Have an honest internal discussion about what “good enough” means for your organization.

3. Is there a packaged application available that meets all criteria, but is difficult to implement because your technology stack is too exotic or idiosyncratic?

Packaged application vendors or a third party may be able to customize an application or pretrained model to suit the challenging requirements of your technology stack. This will cut into time to solution and cost savings, but the customized machine learning component may still be your best option. Enabling new, even if minor, data science initiatives may motivate the organization to consider converting to a more conventional stack and/or updating aging technologies.

4. Is analytics a critical differentiator for your business?

The scale or distinctiveness of your business could mean that you gain greater potential benefits from best-in-class analytics than what your competitors can achieve. If so, building or outsourcing your data science solution is likely the best approach. Buying packaged applications and embedding pretrained models is a good option only when taking on common and relatively straightforward business problems. Businesses that achieve best-in-class machine learning and data science solutions, and that disrupt their industries, typically do so through a “build” strategy or an “outsource” strategy. Consider the sophistication and maturity of the tools involved in a differentiating solution, and how those will impact your need to possibly augment internal staff. Citizen data science tools will be easier to hire or train staff for, while cutting-edge tools will likely require more outsourcing. Well-established open-source tools, libraries and frameworks will be easier to integrate and manage than new, unproven technology. Also consider the life cycle of your solution when outsourcing. Will you be using an apprentice model, where third-party experts help develop the solution then leave it to your staff at the
end of the engagement? Will a third party manage your data science solution indefinitely?

While best-in-class analytics is a tempting and often worthwhile proposition, the potential difficulty of attaining such a solution and the risk of project failure mean that you should also reconsider Question 2 (is there a packaged application or pretrained model available that meets many of — though not all — your requirements?).

5. Is your analytics scenario unique?

If your analytics scenario is truly unique, building a data science solution internally is your best option. In a unique scenario, new and custom-made data science solutions require a great deal of business understanding. Stakeholders need to know why the problem needs to be solved, how the solution will gel with the analytics status quo, and what the future implications of the applied solution will be for the business. Packaged application and API vendors and service providers may not be able to provide the business understanding necessary to make an initiative successful.

Even if your situation is fairly distinctive, however, there is still a chance that a packaged application or pretrained model may be the best solution. If the mechanics of certain business processes in other industries are similar to your business process, you can modify the corresponding packaged application or machine learning service — for example:

- The debt collection business is very similar to database marketing campaigns in retail. A collection strategy (which could include phone calls, SMS or house visits) is comparable to a targeted mail offering that includes a gift, where the cost of the gift may be equivalent to the cost of the particular collection method.

- The mechanics of online gaming or online dating websites mirror the mechanics of certain digital media firms (for example, those that require content-streaming subscriptions).

- Customer churn management shares many similarities with patient care and turnover management in healthcare.

Consequently, many packaged applications and machine learning services can be customized and reused. You should also consider seeking out a data science platform or citizen data science tool with a strong set of precanned solutions for your data science team to adapt.

6. Is your industry one in which new kinds of data have become available, or a new business process has become amenable to data science?

This is a rare situation with enormous upside for organizations with the right analytics mindset. If new kinds of data have become available to you ahead of your industry or domain, reconsider Question 5. In novel analytics situations, it is highly likely that no applicable packaged application or service is available yet, as vendors and service providers may not have had time to acquire any experience in handling this data or process. Weigh the cost and benefits of obtaining first-mover advantage within your industry. Also consider where any new regulatory requirements or internal policies have made customized data science a necessity.

The build option is the most secure for enterprises seeking to win potential first-mover advantage — provided that the contracts of key staff include noncompete clauses. The build option will also help you to avoid unintentionally transferring knowledge to your competitors.

If you are engaging a service provider, the scenario’s uniqueness may warrant entering into a gain-or-risk-sharing pricing arrangement, as the projects in this segment are high-risk and the service provider stands to learn a lot from the engagement. Such
agreements are uncommon, but not unheard of. Typically, companies do not publicly disclose these arrangements.

7. Does your domain require you to be agile?

Agility is highly prized within domains that are subject to rapid change, such as financial markets, social media, or the convergence of operations technology and information technology. Here, data science platforms, coupled with open-source technologies, offer the highest levels of agility and granularity of control. Analytics tasks such as changing up assumptions, or binning customers (specifically, grouping a number of items or values into a smaller number of “bins”) or binning decisions can be accomplished within hours. Open-source solutions without vendor support and citizen data science tools cannot offer the same level of agility and granularity of control.

The build option enables the most rapid change and iteration, as data science teams can tweak and redeploy models to reflect changing business conditions. (Note that in Figure 2, data science platforms offer the greatest flexibility.) The ability to create adaptive models that learn and recalibrate themselves is also available within some data science platforms.

8. Do you have access to data science talent?

If you don’t have any data scientists available in-house, consider hiring good ones whom you can quickly bring up to speed on the particulars of your business (this is easier said than done, but they are out there). If your staff lacks data science skills, de-emphasize data science platforms, open-source solutions and crowdsourcing/talent marketplaces, as all three require a firm grasp of data science and score the lowest for ease of use (see Figure 1). Despite some vendors’ marketing claims of data science democratization, data science platforms are predominantly geared toward users who have solid statistical intuition and knowledge about the principles of data science.

Before you hire pedigreed data scientists or explore outsourcing options, take an inventory of skills among your business analysts and other analytical talent to determine if your organization has potential citizen data scientists (see Note 2). Developers have also become a significant talent resource for data science due to the availability of pretrained models that are packaged as APIs and AI developer toolkits. Also seek out supporting players like data engineers and machine learning operations specialists. Look for individuals with backgrounds in statistics, mathematics, engineering and/or operations research who can be trained through enrollment in massive open online courses (MOOCs) or structured data science training at local universities. These individuals will be valuable assets to the build scenario, provided time to solution is not a priority. For upskilling data professionals in your organization, explore augmented analytics, automated machine learning, and developer-focused functionality and tools. For more information on citizen data scientists and alternatives to hiring traditional data scientists, see “Doing Machine Learning Without (More) Hiring Data Scientists” and “Maximize the Value of Your Data Science Efforts by Empowering Citizen Data Scientists.”

Even if your long-term plan is the build option, consider outsourcing part or all your pilot project to get your data science initiative off the ground. Third-party help will introduce best practices to the organization early on its data science journey and protect it from common pitfalls (such as scope creep or poor data quality). Once you have a data science success story, evangelize its benefits to win support for investments such as full-time data scientists, staff training or dedicated tools. You may also have found a proven partner that you want to work with again in the future, or encountered talented individuals you want to hire full time.
Crowdsourcing and freelancers are often the fastest and most cost-effective option for temporary data science staff augmentation. Crowdsourcing platforms will require (as a minimum) access to knowledgeable staff who are able to appropriately describe the problem setting, the business processes involved, any ambient constraints of the data and a desirable solution. Apart from this, crowdsourcing platforms are the most suitable if the problem is highly self-contained, and the data can be appropriately described and obfuscated or is not highly sensitive.

**Evidence**

The analysis and advice provided in this research are built from constant scanning of the market, as well as from the aggregation of analysts’ experience and ongoing interactions with end users, and technology and service providers. We have used a range of sources to feed our perspective on the topics discussed in this document, including:

- Primary research regarding data science usage and attitudes.
- Gartner client inquiries and conversations with end users and vendors.
- Discussions between Gartner analysts who have expertise in key enabling technologies for analytics and BI, and data science and machine learning.

**Note 1**

**Variations on the Three Solution Paths**

The three solution paths we have described remain the dominant approaches followed by enterprises to obtain data science capabilities, but hybrid approaches are on the rise:

- Collaboration between external and internal teams in building a solution, combining in-house efforts with engaging an external service provider. In this situation, the internal team should drive the process while sourcing in external experts as needed.

- Serware — The constellation in which an external service provider adapts a prepackaged solution that is built for a client’s situation. The client can then further adapt the solution, often working jointly with the service provider, combining a customized packaged solution with the use of an external service provider (see “Take Advantage of the Disruptive Convergence of Analytic Services and Software”).

Other conceivable combinations of the three approaches exist, but are rare.

**Note 2**

**Citizen Data Scientist**

Gartner defines a “citizen data scientist” as a person who creates or generates models that leverage predictive or prescriptive analytics, but whose primary job function is outside of the field of statistics and analytics. The person is not typically a member of an analytics team (for example, an analytics center of excellence) and does not necessarily have a job description that lists analytics as his or her primary role. This person is typically in a line of business, outside of IT and outside of a BI team. However, an IT or BI professional may be a citizen data scientist if that professional’s work on analytics is secondary to his or her primary role.

Citizen data scientists are “power” users who will be able to perform simple and moderately sophisticated analytics applications that would previously have required more expertise. New tools will also make highly skilled data scientists more productive, enabling them to churn out more analysis in the same amount of time. Organizations will still require skilled data scientists for demanding, deep analytics applications. Such data scientists will also have a growing role to play in mentoring citizen data scientists and validating some of their results.

---

*Source: Gartner Research, G00366979, Peter Krensky, Alexander Linden, 30 October 2018*
Interest in big data and machine learning is growing exponentially in many companies and organizations. However, many still feel the great burden of introducing machine learning due to a lack of specialists with required skills and overbearing financial costs.

In order to leverage machine learning, companies must conduct data preprocessing which involves processes such as basic information confirmation, scaling, removal of duplicate values and feature optimization. After they select algorithms for the desired features and build the model, companies must train their model repeatedly until they get their desired forecast results. The chosen machine learning model should then be utilized continuously for prediction on new data. Data scientists and machine learning specialists manually perform all of these processes which can take a very long time to execute. In addition, the machine learning project must be monitored constantly and each model adjusted repeatedly after its’ deployment.

Due to a long period of time required and a lack of specialists, companies cannot easily leverage machine learning. To solve this problem, they need to acquire specialized tools that automate the machine learning process. Considering this, WISEITECH developed WiseProphet with the following three perspectives in mind:

- Can we get a machine learning project off the ground with only data?
- How can we carry out a machine learning project easily and effectively?
How can we support a machine learning project systematically from an engineering point-of-view?

**WiseProphet: A Machine Learning Process Automation Tool**

WiseProphet is a tool that automates the machine learning process, allowing citizen data scientists to perform machine learning tasks. ML projects traditionally require specialists to prepare data, select appropriate variables and models, optimize model hyperparameters, evaluate machine learning models and analyze evaluation results. However, WiseProphet can run various models and set parameters automatically, allowing citizen data scientists to carry out the project without assistance from expert data scientists and machine learning specialists. The following table outlines the four steps of action in the machine learning process.

**WiseProphet Features**

<table>
<thead>
<tr>
<th>Stage</th>
<th>Performance Details</th>
<th>Picture</th>
</tr>
</thead>
</table>
| 1. Understanding the business | ■ Understand business requirements and constraints  
                                  ■ Determine applicable data sources based on business rules |         |
| 2. Data Preprocessing         | ■ Data preprocessing consumes the most amount of time during machine learning projects |         |
|                               | ■ Filter, transform, merge, and group large amounts of data to select important features |         |

Continued on page 14
### 3. Model Training
- Construct a model by applying regression, classification, clustering, and deep learning algorithms to selected features

### 4. Model Operating
- Continuously improve the model by monitoring model predictions

<table>
<thead>
<tr>
<th>Stage</th>
<th>Performance Details</th>
<th>Picture</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Model Training</td>
<td>- Construct a model by applying regression, classification, clustering, and deep learning algorithms to selected features</td>
<td><img src="image1.png" alt="3. Applying Algorithm" /></td>
</tr>
<tr>
<td></td>
<td>- Choose the model that best returns the desired forecasts by repeatedly applying model training and evaluation</td>
<td><img src="image2.png" alt="4. Model Confirmation" /></td>
</tr>
<tr>
<td>4. Model Operating</td>
<td>- Continuously improve the model by monitoring model predictions</td>
<td><img src="image3.png" alt="5. Model Management" /></td>
</tr>
</tbody>
</table>

*Source: Wiseitech*
To automate machine learning, WiseProphet fulfills the following four requirements.

**WiseProphet Advantages**

<table>
<thead>
<tr>
<th>1. Accuracy</th>
<th>2. Special Engineering</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Compare multiple models to choose the best model</td>
<td>✓ Improve performance of data preprocessing</td>
</tr>
<tr>
<td>✓ Train the model repeatedly</td>
<td>✓ Choose optimized features</td>
</tr>
<tr>
<td></td>
<td>✓ Operate tasks easily</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Explanation</td>
<td>4. Operation</td>
</tr>
<tr>
<td>✓ Ensure valid model results</td>
<td>✓ Ensure the created model is distributable</td>
</tr>
<tr>
<td>✓ Ensure transparency of model results</td>
<td>✓ Ensure the model is viable in various environments</td>
</tr>
</tbody>
</table>

**First, WiseProphet makes machine learning**

![WiseProphet](image)

*Source: Wiseitech*
easy to use.

Since the whole process of machine running can be done without having to write code, users without coding experience can analyze data easily.

Understanding Data

![Data Discovery]

Source: Wiseitech

Model Evaluation

![Support Vector Machine]

<table>
<thead>
<tr>
<th>Evaluation Index</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OA_SCORE</td>
<td>0.00</td>
</tr>
<tr>
<td>TEST_OA_SCORE</td>
<td>0.30</td>
</tr>
<tr>
<td>MSE</td>
<td>3.68</td>
</tr>
<tr>
<td>MAE</td>
<td>13.61</td>
</tr>
</tbody>
</table>

Source: Wiseitech
Second, WiseProphet supports both structured and unstructured data. It extracts necessary features from unstructured data such as texts and images and parameterizes them into structured data to be used for machine learning.

![Parameterizing Unstructured Text](source: Wiseitech)

Third, WiseProphet supports both supervised and unsupervised learning. It supports map learning from labeled data as well as unlabeled data using unsupervised learning techniques such as clustering. WiseProphet not only provides basic machine learning algorithms, but also XGBoost, LSTM and deep learning algorithms.

![Unsupervised Learning](source: Wiseitech)
Fourth, WiseProphet provides visualization. The results are presented in the form of a dashboard, allowing a variety of ways to intuitively present data.

Source: Wiseitech
Once the user chooses and deploys the final model, WiseProphet then predicts new data, runs the prediction model via scheduling, and provides management functions to monitor the performance. Also, models created by users through Python can be distributed easily on WiseProphet.

**Monitoring Model**

![Model Prediction Accuracy Chart]

*Source: Wiseitech*

Above all, WiseProphet provides users a means to be more effective at conducting business in their respective fields by bundling industry-specific machine learning models such as predictive maintenance, risk prediction, fraud detection and personalized recommendations.

**WISEITECH’s Machine Learning Methodology**

For companies that need to understand the machine learning process, WISEITECH provides the following machine learning methodology based on the company’s experience in conducting machine learning projects.

1. What do we need to predict?
2. What are the domain concepts?
3. What is the target functionality?
4. Is there a problem with data quality?
5. What measures can we take to increase value of data through data preprocessing?
6. Are there any methods for feature extraction?
7. What kind of model will be used?
8. How do we set the algorithm parameters?
9. What measures do we have in place for data imbalance?
10. Do we have target forecasts from repeated model training?
11. Are there evaluation processes for the prediction model?
12. Are there connections with other enterprise systems?
13. Are there visualizations for monitoring performance?
The following diagram illustrates the machine learning deployment process.

**Machine Learning Deployment Process**

![Diagram of Machine Learning Deployment Process]

*Source: Wiseitech*

**Feature Engineering for Data Preprocessing**

Data preprocessing takes an exorbitant amount of time in order to create a prediction model. Many companies try to adopt machine learning projects and fail because they cannot tolerate the relatively long process of data preprocessing and instead focus too much on algorithms and models. Therefore, it is necessary to understand the business, have a plan for data preprocessing and carefully examine the enterprise’s data quality from the beginning of the machine learning project.
Feature engineering is an engineering approach to data preprocessing that consists of three processes: understanding, improving, and configuring features.

**Three Processes of Feature Engineering**

![Diagram of Feature Engineering processes](image)

**Feature Understanding**
- Data List Generation
- Dictionary Data Generation
- Exploratory Data Analysis
- Selection of Target Variables

**Structured Data**
- Data Transformation
- Cleansing
- Scale Adjustment

**Unstructured Data (Text, Image)**
- Data Transformation
- Cleansing
- Scale Adjustment

**Feature Improvement**
- Feature Extraction
- Feature Selection

**Feature Configuration**
- Feature Extraction
- Feature Selection

*Source: Wiseitech*
**Why do we need feature engineering?**

There are two ways to improve the performance result of prediction. The first is to apply feature engineering and the second is to optimize the hyperparameters of the machine learning algorithm. These are essential tasks that improve the prediction performance but difficult for non-experts to perform.

**Hyperparameter**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>batch_size</td>
<td>128</td>
</tr>
<tr>
<td>epochs</td>
<td>1</td>
</tr>
<tr>
<td>layers</td>
<td></td>
</tr>
<tr>
<td>loss</td>
<td>categorical_crossentropy</td>
</tr>
<tr>
<td>metrics</td>
<td>accuracy</td>
</tr>
<tr>
<td>optimizer</td>
<td>adam</td>
</tr>
<tr>
<td>type</td>
<td>ANN</td>
</tr>
</tbody>
</table>

*Source: Wiseitech*
How much can we improve prediction accuracy when feature engineering is applied?

Prediction Accuracy Before and After Feature Engineering

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>MinMax</th>
<th>Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression</td>
<td>0.97▲</td>
<td>0.96▲</td>
<td>0.96▲</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>0.89▲</td>
<td>0.89▲</td>
<td>0.88▲</td>
</tr>
<tr>
<td>SVM</td>
<td>0.97▲</td>
<td>0.96▲</td>
<td>0.97▲</td>
</tr>
<tr>
<td>MLP</td>
<td>0.96▲</td>
<td>0.97▲</td>
<td>0.95▲</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Standard + PCA</th>
<th>MinMax + PCA</th>
<th>Robust + PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>0.95</td>
<td>0.97▲</td>
<td>0.97▲</td>
</tr>
<tr>
<td>DT</td>
<td>0.88</td>
<td>0.90▲</td>
<td>0.90▲</td>
</tr>
<tr>
<td>SVM</td>
<td>0.97</td>
<td>0.96</td>
<td>0.97</td>
</tr>
<tr>
<td>MLP</td>
<td>0.98▲</td>
<td>0.97</td>
<td>0.96▲</td>
</tr>
</tbody>
</table>

*Prediction accuracy based on original data

Source: Wiseitech

The above example displays result values determining whether a patient has a cancer based on a cancer diagnosis data set. The left side of the screen shows the prediction accuracy when each algorithm is applied without feature engineering based on the original data. The right side shows the prediction accuracy of each algorithm when feature engineering techniques are applied to the data set, with the top right having applied scale adjustment and the bottom right both scale adjustment and principal component analysis (PCA). The result shows that applying feature engineering techniques to the data set improves the overall algorithm accuracy.

Source: Wiseitech
Company Information

Founded in 1990, WISEITECH is an industry-leading big data company specializing in machine learning, big data analysis, data quality and open government data.

It was originally founded under the company name, ‘Wise Information Technology’ in 1990 growing into an industry leader in relational database, data modeling, and data warehousing consulting and became a solution vendor with the development of BI products in 1998.

In 2000, we changed our company name to ‘WISEITECH’ and entered the data analysis solution market and emerged as an industry leader in the domestic professional solution market based on BI, data management and the total solution of CRM.

We are now growing into a company specializing in machine learning and big data company through ‘big data analysis and management using machine learning’.

Contact us

For more information contact us at:

wise.co.kr/eng
contact@wise.co.kr

WISEITECH: Machine Learning Automation Platform “WiseProphet” is published by Wiseitech. Editorial content supplied by Wiseitech is independent of Gartner analysis. All Gartner research is used with Gartner’s permission, and was originally published as part of Gartner’s syndicated research service available to all entitled Gartner clients. © 2019 Gartner, Inc., and/or its affiliates. All rights reserved. The use of Gartner research in this publication does not indicate Gartner’s endorsement of Wiseitech’s products and/or strategies. Reproduction or distribution of this publication in any form without Gartner’s prior written permission is forbidden. The information contained herein has been obtained from sources believed to be reliable. Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. The opinions expressed herein are subject to change without notice. Although Gartner research may include a discussion of related legal issues, Gartner does not provide legal advice or services and its research should not be construed or used as such. Gartner is a public company, and its shareholders may include firms and funds that have financial interests in entities covered in Gartner research. Gartner’s Board of Directors may include senior managers of these firms or funds. Gartner research is produced independently by its research organization without input or influence from these firms, funds or their managers. For further information on the independence and integrity of Gartner research, see “Guiding Principles on Independence and Objectivity” on its website.