Using predictive analytics to improve sales processes and forecasting

Meeting company sales targets and accurately forecasting sales revenue are critical to the success of Microsoft and our customers—and it directly affects the company’s stock price. But it can be hard getting the right insights to more accurately predict the future and take the right actions. To make the jobs of our sellers, sales managers, and sales executives easier, Core Services Engineering (formerly Microsoft IT) is improving the accuracy of sales forecasting. How? We’re using predictive analytics models to better predict the future for sales processes like opportunity management, pipeline management, and forecasting.

By aggregating end-to-end sales information and feeding it into predictive analytics models, our sales teams get powerful analytics-based insights and recommendations. These models are built on Azure Machine Learning and open-source technologies, and are continuously refined based on feedback from sales teams.

The predictive analytics and machine learning algorithms behind the models benefit our sales teams by:

- Increasing accuracy by 98 percent to 99.9 percent for our opportunity ranking model, as machine learning continues to improve and retrain the model.
- Helping sellers and sales managers make data-driven decisions that enable teams to be more efficient.
- Enabling sellers, sales managers, and executives to spend more focused time on customer-facing activities.
- Giving an end-to-end view of the sales processes and desired outcomes.

Our work is part of an approach that we created to modernize our global sales efforts. This approach—Microsoft Sales Experience (MSX)—is built on Microsoft Dynamics CRM Online. It gives our sales teams automated, role-based data, apps, tasks, and actionable insights. We’ve added sophisticated machine learning models to MSX CRM Online to help teams make informed decisions. Instead of replacing the human element of judgment and decision-making, the models augment it by aggregating signals, seller feedback, and intelligence, and then refining the learning over time.

Business challenges

Within the end-to-end sales process, our sellers, sales managers, and executives can face numerous challenges. And what they need often differs from what they’re experiencing.

<table>
<thead>
<tr>
<th>Sellers</th>
<th>Sales managers, sales executives, sales team</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Sellers have lots of customer data, which slows down sales. And bad data’s even worse (garbage in, garbage out).</td>
<td></td>
</tr>
<tr>
<td>• They need a solid pipeline to accurately forecast the sales that they think they’ll land.</td>
<td></td>
</tr>
<tr>
<td>• Sellers don’t feel like they get to spend enough time engaging with customers.</td>
<td></td>
</tr>
<tr>
<td>• Sellers don’t want to waste the time of customers and sales professionals by engaging at the wrong time.</td>
<td></td>
</tr>
<tr>
<td>• They need relevant, accurate insights, for better and faster decision-making.</td>
<td></td>
</tr>
<tr>
<td>• From a personal/professional standpoint, not meeting or exceeding quotas causes stress.</td>
<td></td>
</tr>
<tr>
<td>• Managers need all data in one location in near real-time.</td>
<td></td>
</tr>
<tr>
<td>• Sellers, managers, and executives need to meet or exceed company sales targets.</td>
<td></td>
</tr>
<tr>
<td>• It can be hard staying on track with key performance indicators (KPIs)—for example, close rate (predicting the date when an opportunity will close) and forecast accuracy.</td>
<td></td>
</tr>
<tr>
<td>• The team needs more accurate data to improve forecasting accuracy, and to make better decisions faster.</td>
<td></td>
</tr>
<tr>
<td>• The team needs to reduce human error and subjective judgment to get an objective view of what’s happening</td>
<td></td>
</tr>
</tbody>
</table>
Solutions and models to address these challenges

With these challenges in mind, we created data-driven solutions and models that help our sales teams with opportunity management, pipeline management, and forecasting. From a high-level perspective, we’re creating an easier user experience to assess the present, and using predictive-analytics models to more accurately forecast the future.

New Accelerate Opportunity functionality in Apportal

One solution we’re really excited about is the new Accelerate Opportunity feature. Our sellers use a tool called Apportal, which gives a consolidated view of MSX CRM Online data for relationship management, opportunity management, pipeline management and other areas. To help sellers move sales opportunities more quickly from one stage to the next, we’re adding the Accelerate Opportunity feature to Apportal. It will show all the machine learning that we’re doing for opportunities that involve enterprise customers, along with relevant insights.

Information in Accelerate Opportunity

Accelerate Opportunity has a ton of descriptive, predictive, and prescriptive information.

Figure 1. Information available in Accelerate Opportunity, in Apportal

Guidance on opportunity commitments, based on descriptive, predictive, and prescriptive information

For each opportunity, sellers indicate if it’s committed, uncommitted, uncommitted upside, or committed at risk. And the Opportunity ranking model predicts the probability of win/loss. If there’s any difference between what a seller has committed to and what a model predicts, a Commitment message appears, like the one in Figure 2. But if both the seller and model show that an opportunity’s committed and predicted as a win, the message won’t appear. Eventually, we’ll connect the seller to a coaching framework. But for now, there’s an email link to set up a meeting with a manager.

Figure 2. Prescriptive guidance that combines learnings from Opportunity ranking and Survey Highlights

Survey Highlights and Insights

To give our sales team descriptive information about opportunities, we’re piloting Survey Highlights and Insights within Accelerate Opportunity in selected pilot regions. The information is based on answers to questions in a survey that we give to our sellers.

microsoft.com/itshowcase

January 2017
There are six survey categories, with questions like: What was the competitive advantage/disadvantage? What advice do you have for sellers on similar (same competitor, same product) opportunities? What was the primary factor in winning the deal? Other factors? Impact of Microsoft partners?

We then merged the dataset from the past two to three years, ranked survey comments as high impact and high frequency, created graphs for the six categories, and curated comments related to the opportunity.

**Machine learning–based insights**

In addition to the descriptive survey comments, to show the sales team what contributed to winning a particular opportunity, we plan to include machine learning insights in Accelerate Opportunity. For example, did the customer relationship, pricing, or competitive disadvantage have an impact? And we’ll have related Power BI visualizations.

**Using machine learning models, seller suggestions, predictive analytics**

In Accelerate Opportunity, we’re helping our sales team decide what actions to take. We’re doing this by connecting the output of all models and collecting sales team suggestions, machine suggestions, and prescriptive analytics. To supplement the judgment, expertise, and feedback of our sales team, output appears in Accelerate Opportunity via Survey Highlights, Insights, opportunity ranking/scores (hot, warm, or cold), recommendations, and other information.

Here’s an overview of survey highlights based on seller input, machine learning models, and algorithms that we use.

**Table 2. Machine learning models that we use**

<table>
<thead>
<tr>
<th>Survey highlights</th>
<th>Opportunity ranking</th>
<th>Infopedia</th>
<th>Close date</th>
</tr>
</thead>
<tbody>
<tr>
<td>After an opportunity closes, we distribute a survey to our sellers for opportunities with high customer and dollar value, and that have won or lost against a competitor. Based on six categories of survey questions (like competitive advantage, seller advice, impact of Microsoft partners), we rank responses from the last two to three years and do text analysis. When sellers look at an opportunity in MSX CRM Online, they see insights from peer sellers about how similar opportunities were handled.</td>
<td>Helps with opportunity management and pipeline management. It ranks/scores how likely it is for a seller to win an opportunity—<strong>hot, warm, or cold</strong>. It gives the seller suggestions about how to accelerate moving an opportunity from one stage in the sales cycle to the next.</td>
<td>Retrieves contextual documents from infopedia (an information portal with content for customer engagement), based on the role of the person on the sales team, the account, and the product being sold. This information—which can include compete resources—helps sellers and sales managers better understand the competition and helps them win deals.</td>
<td>Uses historical data to give an objective view, and to more accurately predict when an opportunity will close. The sales manager uses the close-date prediction to see what the seller committed/uncommitted to. Whatever’s committed to goes into the quarterly finance scorecards and projections that affect stock price.</td>
</tr>
</tbody>
</table>

For these models, we use and process algorithms and other input, like:

- Latent semantic analysis (LSA), LSA summarizer, and locality-sensitive hashing.
- Top-ranked statements from the survey, based on closed opportunities.
- Neural network algorithms.
- Deep learning.
- Regression analysis.
High-level flow of our machine learning models

The Opportunity ranking model serves as a template for other machine learning models, which work similarly. The high-level flow works like this:

1. Opportunity management starts in MSX CRM Online. Suppose there’s a new or changed opportunity—like changes to contacts, product-level information, estimated close date, or estimated amount. Changes are pushed from a plug-in from Dynamics CRM Online where they flow into Azure Service Bus. We use two databases:
   - **Sales extended database.** We use this transactional database to store Microsoft-specific account information—for example, customer annual usage of Azure, whether the seller engaged the partner or a competitor was identified. New opportunity scores also go here.
   - **Sales data hub.** We use this for broadcasting and distributing sales data to anyone who wants to subscribe to changes in an opportunity. Older opportunity scores and audit logs also go here.

2. The Predictive Analytics team subscribes to changes in the opportunity and identifies areas for using machine learning. The team uses the models, deciphers the model output, and does telemetry. It also applies business logic and best practices to ensure that results are within threshold values and that there aren’t any duplicates.

3. The Predictive Analytics team uses data services to store the opportunity scores in the sales extended database and maintains history so that we avoid duplicated reports. We give data to the sales data hub for reporting.

4. The data appears in Power BI, MSX CRM Online, Apportal, and chat bots.

Technology used

We manage our sales data with Azure Data Factory, store it in Azure SQL Database, analyze it with Azure Machine Learning, and present the data to users in Power BI, Cortana, MSX CRM, Apportal, and chat bots. Figure 3 shows the high-level technology stack that we use for our predictive analytics sales solutions.

**Technology stack**

*Figure 3. High-level technology stack for our predictive analytics sales solutions*
Architecture for machine learning in sales ecosystem

There are five layers in the reference architecture that we use, with some parts still in development:

- **Training data.** This layer includes our transactional data, outcome history data, sales extended database, sales data hub, data warehouse, and external data sources, like infopedia and complete-related documents. These data sources give sellers and managers additional, contextualized information to help them win deals.

- **Machine learning outcomes.** We calculate model output and store opportunity scores and data in SQL Database, Azure Blob storage, and Azure DocumentDB. Raw, unstructured data input goes in Blob storage and DocumentDB. Then, after we process the raw data with Azure Machine Learning, we store the final output in SQL Database, so that the user interface (UI) layer can read this information as needed.

- **Machine learning models.** Our pregenerated and trained models go here. The machine learning service calls the appropriate model. We can send customized model output/scoring per region.

- **Integration.** Here’s where we have the service bus, message processing, and Azure worker roles. We subscribe to opportunity changes in MSX CRM Online, call the models, and update model scores. We keep integration and UI layers agnostic from models with a service level in between. This way, machine learning services and the UI can “talk” without talking to models, and we can replace older models with updated ones without affecting UI.

- **UI.** Right now, opportunity-related scores/recommendations appear in MSX CRM Online. But when we deploy globally, these will appear in Apportal (a modern app)—in our pilot, sellers already benefit from this. And we’ll have web apps, like a customer planning app and a partner attachment app. We’ll give suggestions to set up a meeting with a customer to get information like the number of licenses they have and to recommend partners for a deal. We’ll also have chat bots that give sellers compete resources for opportunities in real time.

![Figure 4. Architecture for machine learning in our sales ecosystem](microsoft.com/itshowcase)
We created a technical design for machine learning at the worker role and schema level. Here’s how it works:

1. MSX CRM Online contains our transaction information and sales data.
2. When there’s a transaction in MSX CRM Online, data goes to the sales data hub, which is an operational data store with a replica of MSX CRM Online content. Customized attributes go into the sales extended database.
3. We use an opportunity service to abstract the customer data. It gives details based on the opportunity ID.
4. We plan to use Azure Service Fabric to autoscale, based on how much load is generated.
5. Azure Machine Learning models are manually created and uploaded to Azure Machine Learning. We manually process output in Excel, store it in databases, and display it in MSX CRM Online, Apportal, or Power BI. We used Excel just to collect business input on whether our predictions made sense, rather than using it daily in our pilot.
6. The Azure worker role indicates new or changed opportunities. We look for similar opportunities and run multiple models. By using data like win/loss percentage, we offer suggestions about how to win the opportunity.
7. We use different Azure API services to store predictions and suggestions in the sales extended database.
8. The seller goes into MSX CRM Online, can see what the machine predicts, and decide what actions to take.

Increased accuracy and other benefits

Although we’re in an early phase, the benefits from using these solutions internally are already evident. For example:

- The percentage of accuracy as Azure Machine Learning continues to improve the opportunity ranking model is around 98 percent to 99.9 percent, with a goal of 99 percent accuracy or better. This is far better than past predictions that were about 60 to 70 percent accurate. As a result, sales managers and executives have more accuracy for monthly success criteria like forecast accuracy, close rate, and qualified pipeline coverage.
- Azure Machine Learning gave us quick insights from any ad-hoc experiments that we did. For example, suppose a seller created or changed an opportunity in the last 24 hours. With Azure Machine Learning, we run a model as needed, when there’s new or changed information. We don’t have to wait to run it once a month. In some cases, we automated models to run daily, but we can also run them on-demand for retraining and insights.
- From an architectural standpoint, when we run models ad-hoc, we use SQL Database to replicate the data. This way, when we’re retraining the models and recalculating, we don’t use the production database. Instead, we calculate behind the scenes. And when the data’s ready, we put it into the production database. With this design, we can quickly rerun our models on the thousands of opportunities that are updated and created on any day.
- Seller/manager productivity can increase, with less manual effort gathering and tracking information. Our sellers can spend more time talking to customers about sales opportunities, and less time having to get information. Sales managers can spend less time preparing for meetings with their sellers, and more time coaching.
- Data from models helps prioritize opportunities that are predicted to be high revenue and that have a likely chance of winning, along with revealing areas where the model predictions differ from the seller predictions.

Technical challenges and best practices

Some of the technical challenges and related best practices include:

- **Data gathering, categorizing, standardizing, and cleansing.** About a third of our time was spent on these tasks. Just gathering data isn’t enough. We also have to organize and standardize on variations, so that the information is usable and makes sense to our sellers. For example, suppose that we’re competing with another company on an opportunity. To train our models, we restructure our data based on survey categories mentioned earlier—like competitive advantages/disadvantages, seller advice, and primary factors involved in winning an opportunity. We also standardize on terminology, so that sellers have consistent messaging for any opportunity.
- **Data sanity for all models.** For our close date model, there wasn’t enough audit data from MSX CRM Online out of the box. The only way to get this audit data is to configure your CRM Online platform to enable audit logs...
from the start. This way, you can create a close date prediction. In our case, we had to get audit logs from our product group from the past two years just to be able to analyze the data, much less do predictions from it.

- **Size limitations.** Our initial machine learning model exceeded the size limits of our data service. So we split it into several models. The advantage is that the models are more lightweight, so you can rerun them on demand more quickly. So if you have a single model, consider splitting them or creating multiple models from the get-go.

- **Initial issues with training and showing data from a single database.** If training our models is delayed for some reason, the output will be wrong. So we used a secondary database for our calculations. After we’re confident that the data’s correct, we swap the data into production. This takes only about 10 minutes. So if you’re processing millions of records daily, consider using a secondary database for training and then swapping the processed data with production data. This way, the correct data is available quickly in the production system.

- **Initial delays in data availability.** There was a delay of about 30 minutes to move data from MSX CRM Online to our operational data store that’s a replica of MSX CRM Online content. So we implemented a service bus to get the latest data more quickly. As a result, we’ve reduced the delay from 30 minutes to 10 minutes.

- **Piloting and iterating with smaller, role-specific groups first.** We had to iterate and refine a lot so that we could incrementally improve our models. When you create a model, it’s extremely useful to pilot and gather feedback from specific roles and smaller groups, and incorporate this feedback before wider rollout.

- **Increasing seller confidence in model output.** Because we’re still piloting, we’re in the early stages of showing how machine learning models can enhance human judgment and expertise. We show error rates of what models predict versus what happened. If the error rate’s less than 10 percent, sellers will gain confidence over time.

- **Socialize Azure Machine Learning with customers.** We work with customers to see how Azure Machine Learning can solve specific challenges they’re facing.

### Next steps

In the months to come, we look forward to:

- Continuing to socialize Azure Machine Learning company-wide and with customers for specific scenarios.
- Adding automation. We’re focused on visualization now, but we’re starting to build automation into our design.
- Putting our end-to-end customer data (customer 360 data) into Azure Data Lake to enhance our machine learning models and to create churn and cross upseller models.
- Working with internal teams to get customer 360 data like volume licensing and invoicing and build customer 360–related models. This is too cost-prohibitive for just one team to own, so partnering makes sense.
- Continuing to pilot/iterate with smaller groups, before we deploy worldwide.
- Trying to get more analytics and enable logs from out-of-box CRM Online, for more insights on opportunities.

### Summary

Accurate sales forecasting is crucial to the success of our sales professionals and company. Using predictive analytics and machine learning models is an important part of reaching our goal of high accuracy. Continuously improving our models—which we train with accurate, historical data, ongoing feedback, pipeline data, and comparisons of what was forecast versus what happened—are keys to future sales success. We strive to make it easier to predict the future, with the past and present as our guide.
For more information

Microsoft IT Showcase

microsoft.com/ITShowcase

IT Showcase data analytics content

Predictive analytics improves the accuracy of forecasted sales revenue

Using predictive analytics in Microsoft Dynamics CRM Online to improve sales

© 2017 Microsoft Corporation. All rights reserved. Microsoft and Windows are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries. The names of actual companies and products mentioned herein may be the trademarks of their respective owners. This document is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS OR IMPLIED, IN THIS SUMMARY.