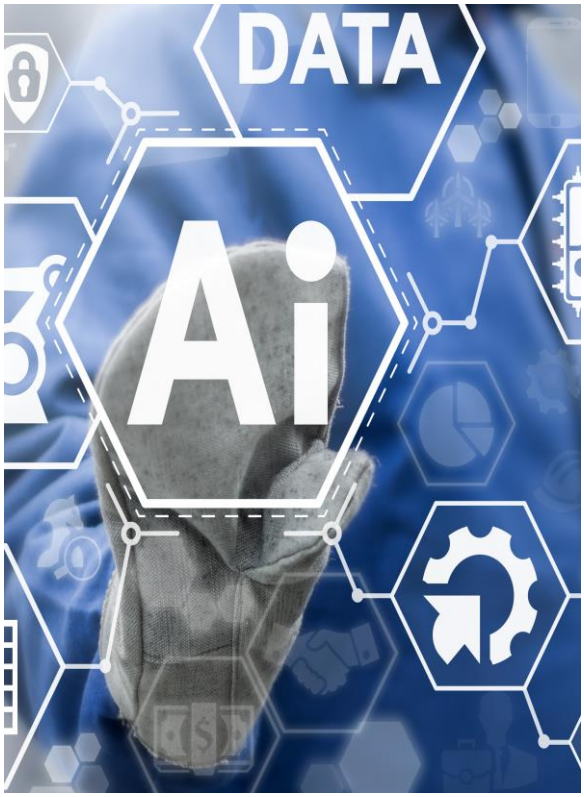



# The promising future of AI (Artificial Intelligence) & ML (Machine Learning)

The Evolution of AI & ML to make data driven decisions and solve business problems using innovative AI systems

ABSTRACT

Learn how Automated Data Science (AutoML + MLOps) can help you make accurate data driven decisions in a couple of days.

**Suman Singh**

CYBORGINTELL

## INTRODUCTION

AI applications are a common place in most industries to help businesses automate, predict, optimize, and innovate. AI and ML are used widely across the world to solve a variety of problems – face and speech recognition, medical diagnosis, self-driven cars, and even robot pets. These advances in technology arrived with the promise of better business decisions and more effective use of resources. Businesses are constantly trying to leverage the power of AI and ML to solve core business problems and make better decisions. The idea of machines augmenting human intellect to analyze, describe and predict data has been profound.

AI, along with the significant backbone of technology and data brings vast value to business. With technology advancing at a pace faster than what it was a decade ago, we have managed to accumulate vast amounts of data and data is what drives AI. Today the volume and velocity at which data is growing poses a huge challenge – harnessing their power in real time or near real time.

Through this thought paper, we will take you through the evolution of AI & ML and its current state brought about by the challenges faced by Business leaders on the effective use of data for faster and outcome-assured decision constrained by tools, skills, and resources.

### But first, what is AI & ML?

Let us start with the basics. What is Artificial Intelligence (AI) and Machine Learning (ML)? How are they different?

In Computer Science, AI stands for any system or program that can perceive its environment, taking in different types of data and then using this data to take actions that maximizes success in a certain task.

Machine learning (ML) is a subset of AI. All ML systems come under AI, but not all AI systems are ML related. The core concept behind ML is that we let machines learn by themselves. Usually, computer programmers write programs to accomplish tasks. However, through ML, scientists aim to avoid 'programming' all knowledge into machines, instead feeding a huge data set to the machine and letting the machine do the learning.

CYBORGINTELL

# THE EVOLUTION OF DATA SCIENCE IN BUSINESSES

***It all started with Data Analysis or Analytics.***

In its nascent stage, it was just Analytics. Data analysis is the process of examining huge data sets to gain insights from them. From scattered sets of data that, by itself, make no sense, a data analyst comes into the picture and tries to find patterns. Through this analysis, they often derive insights and arrive at answers to "what happened?" and "how did it happen?"

***Then came Statistical Modeling.***

With these new demands that data had to meet, it became imperative to deliver more dimensions and scenarios, the need to know what is happening and predict what will happen took centerstage. Statistical modeling takes data analytics one step further. Instead of deriving insights about past events, statistical modeling helped scientists make predictions about future events using the current data. Let us take a simple example - you need to predict over the counter check deposit transaction is risky based on probability of fraud. Statistical modeling will take a data set, find transaction amount more than average balance, or transaction amount greater than historical max transaction amount and can predict transaction which is likely to be fraudulent. This is one of the simplest applications of statistical modeling.

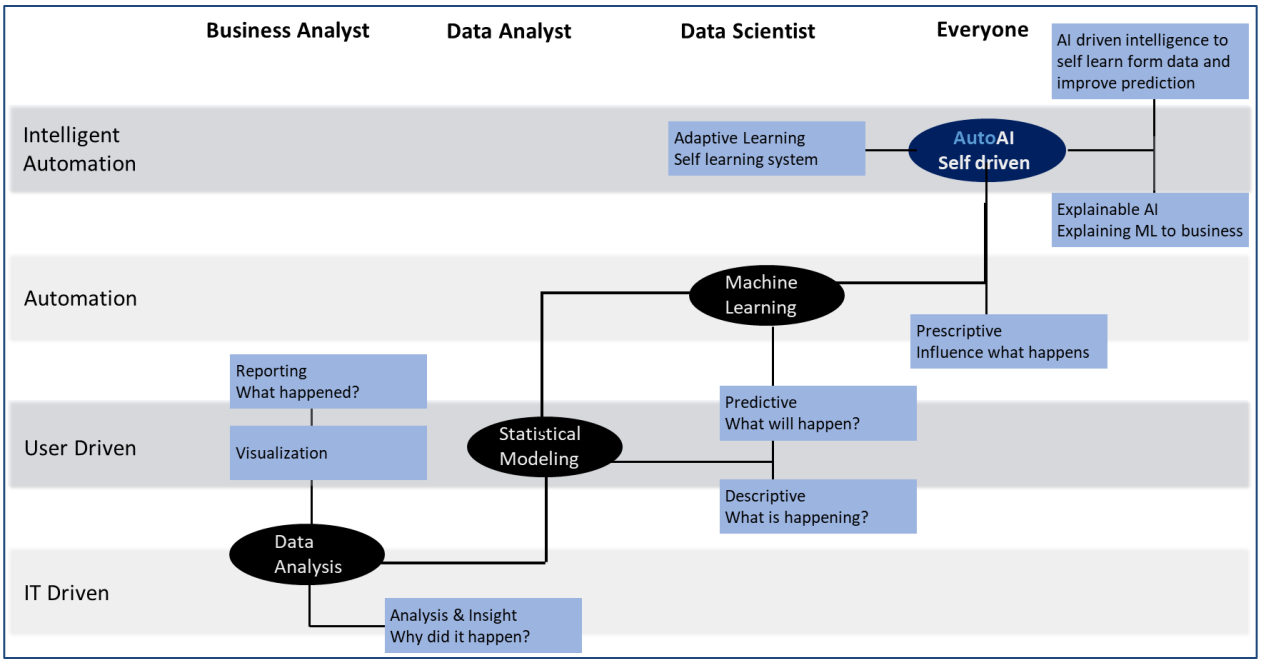
***Machine Learning for prediction.***

Advent of technology tools and availability of huge amounts of data along with the efficiency of cloud and processing capabilities led to the dawn of Machine Learning for predictions. Before machine learning, data sciences and statistical modeling were used exclusively by theoretical scientists and mathematicians. The practical applications and implications were minimal. However, ML has brought in a wave of change in the way organizations make automated decisions. ML has also helped make decisions based on the wisdom acquired from past data. Hence ML influences reactions and decisions and can predict - What will happen? What should happen? Organizations are using more and more automated machine learning models to make quick and informed decisions. These Auto-ML models are so powerful that they can derive knowledge from data and infer from data. **But Auto-ML still cannot solve all of data science's problems**.

The volume of data generated every day is way too enormous for any business to process. Even if a business employs an army of top-notch data scientists and use advanced machine learning models, it would be a mountain of data. For efficient data processing and smart decision making, we need the capability to process vast amounts of meaningful data fast and data scientists must be relieved from the mundane tasks like cleaning the data, taking care of missing values, anomalies and cardinality.

***This heralded the advent of human and machine collaboration: Enter AI Driven Data Science & Machine Learning***

***Exhibit A – Evolution of data science***



CYBORGINTELL

# Automated Machine Learning (AutoML) - AI Driven Data Science & Machine Learning

To enable businesses, the need of the hour is to think one step ahead, to imbibe an adaptive, self-learning system to improve predictions and solutions/actions. Automated Machine Learning (AutoML), as the name suggests, automates the data science/machine learning life cycle (DSML)—data engineering, feature engineering, model development, model evaluation and tracking.

With its unique ability to sift through volumes of information rapidly and objectively, AutoML is becoming the tool of choice when it comes to organizing data in an ever-changing complex world. AutoML enables people with limited knowledge of analytics and machine learning to make informed decisions. It is an even more powerful weapon when you put it in the hand of a data scientist. AutoML has become one of the key weapons of data driven decisioning.

Now let us look at the essential steps in DSML, which AutoML will automate.

**Data engineering**

Data engineering is essentially a preprocessing step that improves the quality of data. AutoML can help a data scientist by automating the essential steps of data engineering like:

- Performing data transformation and validation
- New feature ideation and derivation
- Suggesting the right statistical methods to improve the quality of data: missing imputation, outlier, etc.
- Dealing with large number of categories within a feature using hot encoding, target encoding, WOE transformation, etc.
- Solving the class imbalance problem (class imbalance is when the set of positive data is much larger than the negative set, or vice versa) by:
    - Oversampling
    - Under-sampling
    - Generating synthetic samples using SMOTE (Synthetic Minority Over-Sampling Technique)

- Cohort Modeling method is an approach to improve incidence rate. Generally, we develop multiple modeling cohorts based on the available data. The cohorts are staggered on top of each other to improve incident rates. Please note that due to the staggered cohort design, most customers are included once in each cohort.

**Feature engineering**

Feature engineering is the backbone for developing a highly consistent and accurate machine learning model. A data scientist needs to construct new features (feature derivation) to measure hidden pattern from existing raw data and select the most suitable features (feature selection).

Every data scientist must ask a few important questions before deriving new features:

- Is the list of features suffering from parsimony problem? How can you overcome this issue?
- Can you explain the feature in business language? Can you interpret the feature? Can you measure the impact of the feature on business?
- Do you have methods to track data treatment activities on selected features?
- Are you applying the right feature selection methods? Are you losing important features that can help you to improve the model accuracy?
- Are you using noisy or redundant features in your model, which can lead to overfitting?
- How can you deal with high cardinality data elements in categorical features?

Feature derivation is the key to building a good model. Raw features generally do not result in a well-fitted model, and it also lacks predictive accuracy. AutoML helps data scientists avoid the commonly made mistakes of not knowing how to implement feature derivation. For example, if a particular feature has high cardinality in the data then you need to perform collapsing or indexing.

CYBORGINTELL

## Why Feature Selection

Too many input variables can create complexity while predicting response and result in over fitting. The accuracy of prediction can hamper due to large number of variables in model.

Multiple reasons for feature selection in decision making, few examples are:

- **Better performance**: Avoid inclusion of spurious variables that lead to bad prediction of customer behavior.
- **Limited resources**: Only small number of variables can be collected when enacting trigger in event.
- **Interpretability:** Customer behavior with fewer variables is easier to understand

AutoML ensures that the best features are selected before the model development process. Automation takes away the manual steps and bias involved and ensures high quality models are developed.

## Model development

Model should be developed at the lowest level of detail that is feasible as it can help business to make a more granular decision. Data scientist or Citizen data scientist should have meetings with Business sponsor throughout the model development process to understand the impact of predictors on business, fine tune the list of predictors based on objective. Throughout the model development process, reasonability tests should be applied to predictor variables.

- Time consistency (predictors visualization over a period)
- Cross validation tests (Model development on training and validation on holdout samples)
- Predictors Aging (Predictor's aging is important to maintain the performance of model)
- Contribution of driver across sample over a period
- Data visualization of key predictors with response variable
- Distribution of responses across sample
- Measure over fitting problem if any through cross validation

- Identify cut-off of probability to define sensitivity, Specificity, False Positive & False Negative and compare with prior probability
- In case of regression model, Consistency of regression coefficient and p-value across samples, Consistency of sign of regression coefficient across samples, RMSE across samples

Every data has unique characteristics, which makes it hard for data scientists to know which machine learning model can give better results for the particular use case.

AutoML helps you to automate the entire process of data science machine learning and help you to enable multiple hypothesis and allow you to perform multiple experiments in a matter of few hours to build the best-in-class model. It not only takes away the iterative steps, any many lines of code typically involved in developing models, but also provides you the ability to build several models in parallel so that you can select the model with the best possible results in the shortest amount of time.

**Model evaluation**

Every data scientist wants to ensure that model evaluation is robust. Evaluating a model's performance and being able to trust them is critical for business. However, it is sometimes tricky to finalize model performance based on Area Under the Curve (AUC) or Kolmogorov-Smirnov (KS) chart and similar but limited evaluation techniques. Incorrect evaluation might tamper results, and this, in turn, can lead to business losses. AutoML ensures that these errors do not occur, with their automated capabilities. AutoML ensures several evaluation techniques are applied to your models automatically so that you can be sure of models' accuracy and consistency over time.

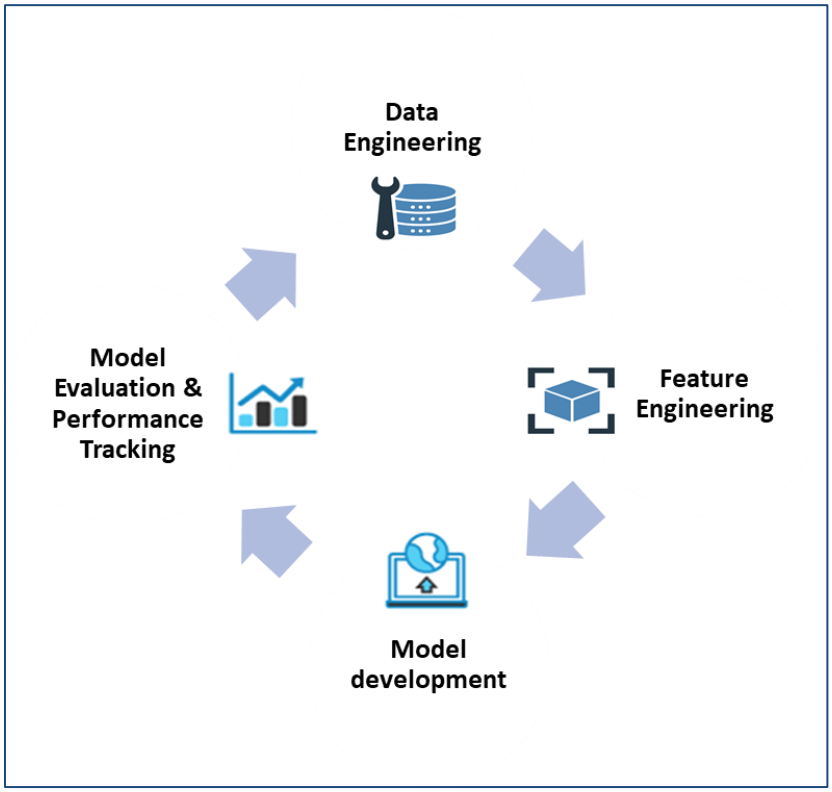
There are other methods to evaluate machine learning models but below are two important model evaluation techniques:

1. **Discrimination** refers to assessing how well your model is able to differentiate between cases in different categories/classes or how much the model can discriminate between cases with value '1' and cases with value '0'. Apart from traditional machine learning model evaluation techniques (Sensitivity, Specificity, ROC, Gini, Lorenz Curve, etc.), data scientist can also measure Briar Score, Precision, Precision AUC, F1, G mean, etc.

2. **Calibration** measures how close the estimates are to a "real" probability. Or how much is the error / discrepancy between the actual and predicted values from the model. Calibration Curves, Goodness of Fit test, Discrimination Slope, etc. provide more power in identifying accurate model.

A robust AutoML solution should allow you to measure model performance and evaluate it automatically and using several novel techniques, including those described above. This ensures that data scientists spend time on actually generating business value, and also give them the necessary confidence to defend their model's performance with business users.

***Exhibit B – Auto ML***




**Most AI & ML solutions stop here – just automation of the traditional DSML life cycle of data engineering, feature engineering, model development and model evaluation.**

However, as per Venture Beats - Transform 2019 Report, 87% of Data Science and Machine Learning projects do not progress beyond prototype, research & development stage.

There are various reasons for failure of AI model in production like wrong feature derivation while building model, challenges in data processing, but more importantly the difficulty in deployment, lack of Explanability and simply the lack of trust in the model's results.

To accelerate AI & ML adoption, solutions need to go beyond model development. This includes Explainable AI, Automated Model Deployment, Auto Performance Monitoring, and Continuous learning.

To ensure AI models move from the lab and into production, and to ensure lack of business impact due to AI failure, businesses need more than just AutoML, they need **MLOPs** to ensure that machine learning projects can be easily and automatically deployed in complex production environments. Predictive accuracy and business value needs to be measured and improved on an ongoing basis and adapt to constantly changing conditions, ensuring you get the ROI you are looking for.

The below capabilities should be available in an MLOps platform to ensure that the power of AI stretches beyond AutoML.

1. **Model Implementation** to automatically deploy ML & DL models across libraries in diverse production environments. Businesses need the ability of real time scoring using streaming data functionalities, automated feature derivation pipeline and real time scoring in a few milli-seconds. Most businesses have been on an AI journey for a while, this means they have probably built models with different tools over time. They need a single statement of record for all AI & ML models in the organization. They also need to easily integrate AI models with existing systems in legacy architectures with Rest APIs that can share both predictive scores and drivers of these predictions for real time decision making and customer actions.
2. **Model Management & Maintenance:** The ability to configure machine learning models and defining the model versioning to track models is key. MLOps should allow users to govern models in production such as swapping shadow and live

models, deleting poorly performing models. MLOps should include approval workflows for managing models in production. Executive Dashboard of AI in action provide visibility of model performance in production. These enable you to know how many observations have been scored, model health status, etc.

3. **Model Performance Tracking:** Model performance tracking in production is one of the most important aspects in ensuring adoption of data science machine learning. It allows business and IT to ensure that every prediction is accurate and consistent. Poor performing models in production, and failure of AI models can damage the business gravely. To avoid any business impact, you need the capability to monitor success and failure. MLOps tools should consider this and allow users to measure the risks and rewards of machine learning models.
4. **Model Audit Trail System:** Transparency and Explanability of AI is critical for business adoption and trust of data science and AI models. Data scientists need detailed explanations of models, results, predictions, and drivers of prediction both at model and record level. They would require automated documentation of the entire data science process and enables audit of decisions, codes, processes, and methods, which are critical to meet compliance and regulatory requirements, but most importantly they need to make business users to begin to trust AI driven decisions.

Automated data science machine learning can help businesses to accelerate their AI journey and increase adoption of predictive intelligence driven decisioning. Earlier to solve a business problem with Data Science, it would require a team of data scientists, data engineers, technologists, and business experts and would take a few months to develop and operationalize AI model.

In comparison now, AutoML+ and MLOps enables businesses to reduce the time from data to decision to a few days. The power of AI driven technology is not just speed and accuracy but also allows the data scientist and business user to collaborate more with stakeholders and focus on building innovative solutions to business problems instead of performing repetitive, redundant and manual tasks in the data science lifecycle.

CYBORGINTELL

# iTuring

## Our innovative solution to completely automate the DSML lifecycle.

iTuring AutoAI from CYBORGINTELL is a revolutionary product built by data scientists and AI experts with years of research. We have gone deep into the issues faced by businesses and data scientists and used this knowledge and innovation to build iTuring AutoAI, which is a next-gen system that automates the DSML life cycle.

**iTuring AutoML+** is a zero code, intuitive UI democratizes data science and puts the power of data science and machine learning in the hand of your employees without worrying about the complexity. By automating the entire lifecycle – data transformation, feature engineering, model development and evaluation, data scientists now have the capacity to address more business problems faster and at scale. With detailed explanations at model and transaction level, iTuring provides data driven insights in hours instead of weeks for business leaders to make informed decisions.

With **iTuring MLOps**, machine learning projects can be easily and automatically deployed in complex production environments in a few clicks, irrespective of where they were built. Predictive accuracy and business value is measured and improved on an ongoing basis and using CI's Dynamic AI, models adapt to constantly changing conditions, ensuring you get the ROI you are looking for. Real time scoring and production ready rest APIs ensure an automated decision-making capability can be integrated into a company's operations to automate decisions at rapid scale.

**To learn more about our flagship product iTuring, please visit www.cyborgintell.com.**