Machine Learning for Business Value

You know Machine Learning (ML) and AI can deliver value. A quick chat with an AI agent or an online search can generate a long list of plausible ML projects in seconds, while scores of case studies with hyperbolic ROI published daily keep expanding the options. The challenge is to shortlist the right candidates for your business, which requires methodical screening to discern signals amid the noise. An ML project that delivers business value meets two critical requirements: feasibility and business impact.

1. Feasibility

An infeasible project is bound to stall. ML success requires business expertise and carefully curated data to map relationships between various features (parameters) and the target within regulatory frameworks. Therefore, it is critical to validate that a sufficiently stable and predictive relationship exists, that adequate data is available and permissible under regulatory and legal frameworks, and that essential resources will be available.

1.1. Real Stable Relationship

ML works by extracting patterns from historical data and using them to predict the outcomes, which only works when a genuine and sufficiently stable relationship exists between the inputs and the outcome in the real world. Any attempt to build an ML model without such a relationship will either stall the project or produce a fatally inaccurate model.

An effective screening test is: Can a domain (business) expert predict the outcome with reasonable accuracy given a set of input data? If the answer is no, it requires revisiting the objective and approach.

1.2. Data Feasibility

Where a real-world relationship exists, ML needs meaningful data to learn it and generate predictions. Data feasibility rests on three pillars: the availability of meaningful data, the accessibility and usability of these data for training, and the timeliness of essential data for generating predictions.

1.2.1. Meaningful Data Availability

The data availability requirement expands beyond data volume. Although data volume is a critical component, various factors are essential for data to be meaningful for ML modeling, such as:

- **Feature Availability:** The target value depends on a range of factors, and accurate prediction requires data representing each factor. The absence of a single feature does not automatically render an approach infeasible; its impact depends on how strongly the feature influences the target. However, feature absence does affect model accuracy. Sometimes a proxy feature is required when a direct measure is unviable, such as measuring emotion, or is extremely expensive to record, such as real-time tyre tread wear.

- **Data Volume:** The volume of data required to teach the model the relationship in the desired range of input variables depends upon the number of input features, relationship complexity, and the desired range, including rare but important cases.

- **Feature Variation:** The mere presence of a feature isn't enough, as a fixed unit price over a long period can only teach the model how sales vary under different economic conditions at a particular price, not price elasticity, let alone how elasticity shifts across market conditions. Therefore, sufficient variation across the relevant conditions is equally essential.

- **Granularity:** The available data must match the granularity of the intended prediction; for example, hourly sales data should be present to predict hourly sales. Finer granularity data can be aggregated to the desired level.

1.2.2. Data Usability

Data availability does not automatically translate into feasible usage for modeling. The following criteria need to be met to achieve usability:

- **Connectivity:** Data may be available with the detail and depth necessary but fragmented across systems, with no way to join them accurately to generate the necessary training dataset; for example, customer and sales data may be present, but mapping each customer to their purchase history is implausible, rendering any plan to generate product-specific recommendations infeasible.

- **Quality:** Data must be sufficiently accurate, complete, consistent, and representative for the intended use. Missing values, incorrect records, inconsistent definitions, measurement errors, and changes in data-collection practices can undermine an ML model's ability to learn a reliable relationship. Many techniques exist for cleansing data, but lower-quality raw data degrades accuracy, and large inaccuracies across multiple features or systematic errors may reduce accuracy below the acceptable threshold.

- **Data Usage Regulations:** The regulatory landscape for ML data varies by jurisdiction, but rests primarily on two principles: Data Privacy and Anti-discrimination. Data privacy frameworks govern personal data through core tenets including collection, purpose, and use limitations. Anti-discrimination legislation prohibits discrimination based on traits such as age, religion, sex, race, ethnicity, nationality, sexual orientation, disability, and political opinion, depending on the jurisdiction. These frameworks can prevent highly informative features or useful data from being used.

1.2.3. Data Timeliness

Including a feature in a model that won't be available at prediction time can lead to a fundamental mistake, discovered only after the project is complete. The timeline of data availability is often overlooked during training, as it is performed on batches of historical data. The feature used in development may not be present during prediction because of:

- **Post-Prediction Generation:** A feature may not be available because it is generated after the prediction requirement; for example, using the number of claims by customer to predict the propensity to commit insurance fraud, which won't be available at the time of onboarding.

- **Capture Frequency Mismatch:** Data may exist but be captured at different frequencies; for example, a restaurant franchisor may record individual invoices across locations but only fetch all data at the end of the day; thus, they won't have timely data to adjust hourly offers based on sales.

1.3. Regulatory Feasibility

A technically viable model supported by adequate data may still be infeasible because of legal and regulatory constraints. AI-specific laws and industry regulations can restrict the applications, how models are developed, and how predictions can be used.

1.3.1. Prohibited Applications

A few ML applications may be outright prohibited. These prohibitions can come from AI laws, such as individual predictive policing based solely on profiling under the EU AI Act, or industry-specific laws, such as using genetic information to price or deny health coverage under the U.S. Genetic Information Nondiscrimination Act.

1.3.2. Compliance Burden

Many applications become unviable because of the range of requirements imposed on their development and use. These may include restrictions on training data and methods; requirements for explainability and documentation; extensive testing and validation; human oversight; ongoing monitoring; and regulatory approval before deployment. For example, a requirement for a sufficiently explainable model may rule out an otherwise suitable ML architecture when its complexity makes the required level of explanation impractical. Similarly, validation and approval requirements may make an application economically or operationally unsound even when the underlying model performs well.

Regulatory feasibility requires assessing whether an application is permitted and, where permitted, whether the associated compliance requirements are compatible with the proposed ML approach and business case.

1.4. Expertise and Resources

Even a scientifically and legally sound ML application can stall without essential tools and expertise. The exact requirements vary with the chosen approach to model development and deployment.

1.4.1. Tooling & Infrastructure

Resource feasibility spans four areas: infrastructure for data storage and processing, compute and training hardware, compliance readiness tools, and environments and tools for model deployment and prediction use. The mere absence of a tool should not make a potential ML application ineligible. However, if acquiring or developing the required capability takes longer than the project can accommodate, or requires resources beyond what the business can commit, the timeline or approach needs to be reevaluated.

1.4.2. Talent Gap

The pathway chosen to build ML models dramatically influences the expertise requirement, varying from requiring a large team of specialists involving Data Scientists, MLOps Engineers, Data Engineers, Software Engineers, among others, to leaning on data and domain expertise already within the business. Expertise gaps are generally addressable through hiring, training, or external support; thus, the key constraints are often the time available and the level of business commitment required before results can be delivered.

2. Business Impact

Businesses operate to deliver exceptional value to their customers and make a profit in return, not simply to build AI models. An ML project without clear business impact is an expensive scientific experiment. ML delivers business impact when its predictions lead to actions that create measurable value that justifies the investment required to build and operate the solution.

2.1. Actionable Prediction

A model can be accurate yet deliver zero return if no team or system acts upon its predictions. Therefore, the prediction must trigger or alter an action, which requires it to be workflow native and earn the team’s trust to act on it.

2.1.1. Workflow Native

A prediction that requires hiring a new team or substantially restructuring operational processes often fails because of organizational friction and cost exceeding potential financial return. Therefore, predictions must integrate seamlessly into existing workflows.

- **Execution Capability:** An ML model predicting customer churn likelihood generates no value if no team engages customers and no automated system delivers a retention offer.

- **Decision-Ready Prediction:** A model predicting a customer's propensity to purchase given a discount provides limited value if the operational decision is how much discount to offer.

The model should therefore generate outputs that map directly to existing operational decisions. The closer a prediction is to an existing decision point, the less organizational effort is required to convert model output into action and, consequently, the greater the likelihood of realizing business value.

2.1.2. Team Trust

A model that lacks the team’s trust gets circumvented, irrespective of its theoretical accuracy. Trust cannot be built overnight; it develops with each success. Nevertheless, a few ingredients are critical to gain initial trust for its adoption:

- **Predictive Visibility:** Users need sufficient visibility into the factors influencing a prediction to assess whether it makes business sense.

- **Boundary Clarity & Safeguards:** Teams need clarity around situations where the model may be unreliable and how the solution automatically isolates those data points.

- **Business Metrics:** Relevant business metrics need to be employed alongside technical accuracy metrics for evaluating and documenting model performance.

A model that is technically accurate but routinely ignored or overridden by the team is unlikely to deliver its intended business value.

2.2. Net Financial Return

The relevant question is not whether ML can generate value, but whether it generates enough incremental value over the best practical alternative to justify its total cost and risk.

2.2.1. Measured Gain

The financial gain from an ML model is challenging to compute, especially before building it, as its accuracy directly influences the benefit. The additional factors specific to evaluating the financial return of ML models are:

- **Heuristic Baselines:** It’s tempting to reach for ML because of the hype, but often a human expert’s judgment encoded as a handful of conditional rules captures most of the variation. Therefore, it is essential to consider the incremental return from easy-to-build and maintain heuristic models.

- **Model Accuracy:** No model is completely accurate; therefore, return must be computed based on plausible model accuracy, accounting for lost opportunities.

- **Cost of Inaccurate Predictions:** Beyond missed opportunities, inaccurate predictions can also create direct costs, depending on the application. A false-positive churn prediction leads to offering retention promotions to engaged customers, reducing margin.

- **Ability to Act:** Prediction volume does not equal value-generating action volume. If the team and system are constrained to act on only a handful of predictions, then financial return is bound by that, regardless of the number of predicted data points. A team that can only call 100 customers a day can act on only 100 potential churning customers, no matter how many are predicted.

- **Legal & Liability Exposure:** An inaccurate prediction without ample guardrails could lead to large settlements, particularly when it directly impacts someone's well-being, such as an incorrect structural safety audit or medical diagnosis.

2.2.2. Total Cost of Ownership

A common pitfall is considering only training time and resources while neglecting deployment, which consumes significant human resources, and prediction, where everything from monitoring to acting on generated predictions requires continuous expenditure. The primary drivers of the cost are:

- **Human Capital:** It is a primary cost driver in most ML projects. The required expertise and time for development, deployment, monitoring, and maintenance depend on the approach and tools. The range extends from a large team of different specialists spending multiple quarters building a model in-house using open-source tools to a single individual overseeing an outsourced project. The expertise and time required to deploy models are often underestimated, along with the time spent troubleshooting pipelines in production.

- **Infrastructure and Tools:** Infrastructure cost estimates need to expand beyond compute expenses to include data storage and the tools necessary for training, deployment, compliance readiness, monitoring, and generating predictions.

- **Regulatory Compliance Cost:** Every ML application has basic regulatory and legal requirements, such as documentation, internal validation records, and audit logs, which require initial investment and ongoing maintenance. Furthermore, a host of application and domain-specific compliance requirements may apply, such as additional rigorous testing, regulatory approval before deployment, detailed documentation, and human-in-the-loop decision-making. These additional compliance requirements could significantly alter the timeline and total cost.

- **System and Team Acting on Prediction:** Translating predictions into outcomes may require changes to existing software and workflows, recruiting a team, building API connections, and modifying interfaces. Downstream changes are often overlooked until the operationalization stage, such as the need to transition to digital menu boards at all restaurant locations to implement hourly dynamic pricing, which can disrupt the anticipated ROI at the final stage.

- **External Services and Data:** These can easily inflate or shrink the budget depending on how and stage they are employed. A common temptation is to exclude outsourced data collection and curation costs, arguing that the work could be reused across multiple projects. However, different problems may require different data and preprocessing, and data may drift, making current efforts largely irrelevant to future projects. Therefore, every cost associated with the project, from data collection and preprocessing to development and prediction, needs to be accounted for in evaluating business feasibility.

- **Reputation and Social Risk Mitigation:** Businesses increasingly face scrutiny over the use of personal data and AI in decision-making. Therefore, reputation risk remains even when operating within a legal and regulatory framework. These risks can be particularly high when ML influences decisions affecting fundamental aspects of people's lives, such as employment through resume screening and AI interviewing, food through dynamic pricing and user-specific pricing, or freedom through automated surveillance or tracking. Therefore, it is critical to evaluate whether the proposed ML project could face public backlash and compare the potential costs against the expected return. Clear, transparent communication and stakeholder engagement could mitigate much of the anticipated backlash, but the evaluation and budget should include their costs.

Conclusion

Machine learning creates business value not by producing accurate predictions alone, but by solving a feasible business problem and translating those predictions into measurable outcomes. Before investing in an ML project, businesses should therefore assess both feasibility and business impact.

Feasibility requires more than confirming that data exists. A viable application needs a sufficiently stable relationship between inputs and the target, meaningful and usable data with adequate variation and appropriate granularity, timely access to essential information, a permissible regulatory environment, and the expertise, infrastructure, and resources required to develop and operate the solution.

Business impact requires an equally rigorous assessment. Predictions must lead to actions that fit existing workflows and can be trusted by the teams using them. The resulting incremental value must then be sufficient to justify the total cost and risk of the solution, including model errors, operational changes, compliance requirements, ongoing maintenance, and potential legal, reputational, and social consequences. Simpler alternatives, including human judgment and heuristic rules, should be considered when evaluating the incremental value of ML.

Ultimately, the right question is not whether a business can build an ML model, but whether it should build one. A strong ML opportunity exists where a real and sufficiently stable relationship can be learned from feasible data, the resulting prediction can drive an actionable decision, and the incremental value of doing so exceeds the full cost and risk of delivering it. Applying these principles systematically can help businesses focus their ML investments on applications with a credible path from data and prediction to sustained business value.