Axilient

Enterprise AI, ML and Data Science Operationalization

10 Things You Should Consider Before Choosing a Data Science Partner

Overview

Artificial Intelligence (AI) has become an integral part of our society, whether it be a headline in the news or its practical application in our everyday lives. AI has been able to do some truly remarkable things like recommending the next television show you want to binge watch, assisting doctors in robotic surgery, and flagging fraudulent credit card transactions. Across the board, AI has reduced costs, improved productivity, created efficiencies, and increased revenue.

Overview

Artificial Intelligence (AI) has become an integral part of our society, whether it be a headline in the news or its practical application in our everyday lives. AI has been able to do some truly remarkable things like recommending the next television show you want to binge watch, assisting doctors in robotic surgery, and flagging fraudulent credit card transactions. Across the board, AI has reduced costs, improved productivity, created efficiencies, and increased revenue.

Artificial Intelligence

Artificial Intelligence is the ability for computers to perform functions that would normally require human intelligence. It is a process and not something that you can tangibly see or feel. AI is comprised of several competencies, the most common being Machine Learning (ML), which uses data to power complex algorithms to make correlations that drive business decisions.

Netflix is a household name and service that is almost entirely based on AI. Their streaming service has been trained to recommend content rather than having consumers make a choice about what to watch. It is a prime example of a computer performing a task generally assigned to humans. Algorithms use data from millions of subscribers to determine exactly what each person wants to watch. This hyper-personalized structure is ideal in an exponentially growing digital world.

Consumer Data

Businesses have access to large amounts of consumer and product information that takes on many forms including transactional, demographic, and behavioural data. With access to so much data, virtually every business has the potential to make gains using AI. While you might not become an overnight AI pioneer, there is great potential to be realized when new ideas are embraced.

While it may seem like a no-brainer to invest in AI, there are still various factors that hold companies back:

- AI is dismissed as hype that will not be able to deliver results
- Leaders are unsure how AI can be applied to their business
- The financial and technical barriers seem too high to surmount

AI Vendors

Truth be told, very few companies have the skills, technology, or foresight to make use of AI on their own making it critical to find the right AI vendor for your business.

It would not be feasible for most companies to develop their own AI team because their goals are often undefined, the costs are high, and recruiting the appropriate talent is extremely difficult. This coupled with the growth potential of AI has made artificial intelligence vendors increasingly popular.

There are a large number of technology vendors anxious to help the next company that wants to invest in artificial intelligence and machine learning. According to Crunchbase, there are now over 8,000 vendors in the AI marketplace. These vendors specialize in data science platforms, machine learning capabilities, and various tools that can optimize your business and propel it into the world of AI. Having such a large number of vendors makes it very difficult to identify the best fit for your business.

Vendors fit into many categories and range from broad AI service providers to industry specific solutions. For example, some vendors will specialise in developing AI for healthcare or retail while others focus on emerging technologies such as autonomous vehicles. Whether you need a vendor that has a strong understanding of your industry versus one that takes a broad stroke to AI depends on how well-defined your problem is and where your company is in the AI lifecycle.

Consumer Data

Businesses have access to large amounts of consumer and product information that takes on many forms including transactional, demographic, and behavioural data. With access to so much data, virtually every business has the potential to make gains using AI. While you might not become an overnight AI pioneer, there is great potential to be realized when new ideas are embraced.

While it may seem like a no-brainer to invest in AI, there are still various factors that hold companies back:

- AI is dismissed as hype that will not be able to deliver results
- Leaders are unsure how AI can be applied to their business
- The financial and technical barriers seem too high to surmount

Truth be told, very few companies have the skills, technology, or foresight to make use of AI on their own making it critical to find the right AI vendor for your business.

Artificial Intelligence Vendors

It would not be feasible for most companies to develop their own AI team because their goals are often undefined, the costs are high, and recruiting the appropriate talent is extremely difficult. This coupled with the growth potential of AI has made artificial intelligence vendors increasingly popular.

There are a large number of technology vendors anxious to help the next company that wants to invest in artificial intelligence and machine learning. According to Crunchbase, there are now over 8,000 vendors in the AI marketplace. These vendors specialize in data science platforms, machine learning capabilities, and various tools that can optimize your business and propel it into the world of AI. Having such a large number of vendors makes it very difficult to identify the best fit for your business.

Artificial Intelligence Vendors

It would not be feasible for most companies to develop their own AI team because their goals are often undefined, the costs are high, and recruiting the appropriate talent is extremely difficult. This coupled with the growth potential of AI has made artificial intelligence vendors increasingly popular.

There are a large number of technology vendors anxious to help the next company that wants to invest in artificial intelligence and machine learning. According to Crunchbase, there are now over 8,000 vendors in the AI marketplace. These vendors specialize in data science platforms, machine learning capabilities, and various tools that can optimize your business and propel it into the world of AI. Having such a large number of vendors makes it very difficult to identify the best fit for your business.

Vendors fit into many categories and range from broad AI service providers to industry specific solutions. For example, some vendors will specialise in developing AI for healthcare or retail while others focus on emerging technologies such as autonomous vehicles. Whether you need a vendor that has a strong understanding of your industry versus one that takes a broad stroke to AI depends on how well-defined your problem is and where your company is in the AI lifecycle.

Underwriting and Pricing

Underwriting requires large amounts of human resources thus it does make business sense for introducing machine learning in insurance. For the year 2017, USA had nearly 350 billion of P/C losses. These numbers continued to rise at a much higher rate since 2013 when compared to the growth of premium. There was some improvement in 2018 with better premium growth along with stagnation of the loss growth but the situation remains dire due to higher losses overall.

P/C insurers have welcomed machine learning models with open arms as they are its greatest proponents in recent history. Consequently, life insurers have shown a keen interest in machine learning as well. While they are able to get accurate figures for mortality rates from the actuarial tables, the challenge lies in modelling the persistence or ability of the customer to make payments over their lifetime. This situation is tailor made for the use of machine learning models.

Recent industry trends call for an improvement in the efficiency of underwriting. It is expected that insurance needs will become more personalized with the passage of time. This means that there would be individual expectations at different times for different purposes. According to McKinsey, consumers will rely on discrete purchases which can be renewed annually. This will create a continuous cycle of seamless interactions as products with micro-durations are bought more frequently with greater demand.

Similarly, it is also foreseen that insurance products will be broken down into small and specific segments (for e.g. phone battery insurance, TV insurance etc.) as a consequence of reduced costs and time required for purchases. It presents an opportunity for insurance companies to make money in the same vein as US airlines who introduced additional charges for specific services. This may increase the workload in terms of underwriting and could lead to greater expense if more resources are dedicated to it. Machine learning can minimize underwriting expenses over a period of time.

The question arises how machine learning will cut down the expenses incurred from underwriting activities? Well, broadly speaking, it can achieve this target in two ways. The first one would be the facilitation in automating the underwriting process. As per a study conducted in 2017, around half of the companies are on track to automate their underwriting work. This means that marginal cost for every additional customer will drop and, as a result, insurers can enjoy more profits.

Such a scenario is more likely to occur if the volume of the customers is high. The second way through which machine learning can help insurers is through better handling and organization of unstructured data. By sorting the data, a lesser human effort is required to make sense of it. It may even uncover data that was present within the company but remained unused so far. More meaningful information can be extracted from individual data points which, in turn, leads to a better return from investing in a machine learning model.

There are two recent development which are increasing the effectiveness of machine learning models in the domain of insurance. The first technology, Natural Language Understanding (NLU), allows insurers to sift through data from non-conventional sources such as LinkedIn, Facebook etc. As a result, information from these sources is also factored in while making decisions.

The second technological development is the growth of IoT (Internet of Things). Insurers are able to leverage IoT to get the data directly from customers instead of relying on indirect sources. As a result, insures are able to make use of usage based insurance (UBI) by gathering customer data from various devices. This information is used to minimize customer risks by finding the right pricing and application of their insurance.

With the UBI approach, insurers can mitigate problems by notifying the customers beforehand. UBI can prove to be successful as customers are generally more open to sharing information to insurance companies in comparison to other sectors. As per the statistics, nearly 50 % of the customers were open to sharing their information for auto insurance (about 40% were unwilling). Similarly, around 40 % were willing to share information about their homes with a similar proportion being unwilling to do so. It was observed that customer's willingness to share data was dependent on the fact that whether the choice was given to opt out of sharing data for UBI.

Older models restricted the amount of data used to underwrite customers to a few discrete variables. Machine learning has broadened the horizons by not only presenting a larger set of variables but by making new links between previously unrelated data. In the past, traditional regression models may use the customer's current address to draw conclusions during underwriting. With machine learning, along with the current address, the history of the customer's previous addresses is factored in while making decisions. Consequently, machine learning models can figure out the determinants for the likelihood of a customer filing a claim by combining different data points.

These new technologies are facilitating an unparalleled level of in-depth analysis of data. This has led to improvement of the rates set across different parameters of the premium which includes risk relativity score, the rate change and base rate. Rate values are more customized and tailor made in accordance to factors like region, state and industry. Additionally, machine learning models have made it easier to link the impact of economic conditions on the rates which can, in turn, be used to predict the possible risks.

Servicing and Claims

Methods of pricing and servicing claims have evolved with the introduction of machine learning. It can observed in the market that early adopters of the ML model are able to predict the average number of claims along with the costs associated to them. This puts companies in a much better and stable position. They are able to ensure that enough funds are available in the reserves to pay off any pending claims. Moreover, prices of different products can be adjusted to match the risk associated with the policyholder.

In the servicing department, machine learning is minimizing human involvement during interactions with the customers. In fact as per the estimations by Gartner, it is expected that more than 80% of the customer dealings with industries would be done without humans.

There are examples available in the market where AI has replaced humans during customer to business interactions. Lemonade, a property management insurer, has introduced chatbots which are capable of handling customer queries and simple claims. Another example is of Fokuku Mutual Life Insurance. In December 2016, they announced that they will lay off 30% of their customer service staff and replace them with AI. Initially, the system will cost around 2 million dollars to implement but will lead to annual savings of over 1 million dollars in the form of reductions in salaries payable.

Fraud

Fraud cases have been a thorn in the side of the insurance companies. The Insurance Information Institute estimates that there are nearly $ 35 billion P/C losses due to fraud. There are different sources of frauds, some within companies like service providers for claims while others are outsiders like policy holders and applicants.

Insurance fraud falls under two categories: hard and soft. Hard fraud involves forgery of documents and falsified information such as staged accidents. While on the other hand, soft fraud mixes facts with lies in order to inflate the claim value. This could be done by over-reporting losses or by under-reporting number of employees. There are certain areas in P/C such as employee compensation and auto policies which are most susceptible to fraud.

Machine learning can a valuable tool for minimizing fraud losses. It can minimize premium leakage during underwriting where incorrect or incomplete information leads to skewed assessments. As per estimates of One Verisk, nearly $ 30 billion were lost due to premium leakage in automotive insurances. Machine learning can help in recognizing potential fraud threats by categorizing customers on the basis of their risk factors. It can also conduct additional sweeps to double check for any new fraud indicators within the customer base.

Challenges to Adoption

Regulations

Regulatory control of machine learning is important. It ensures that the customer base is not exploited and there is no discrimination of any particular group. Therefore, companies need to plan their decisions accordingly.

There is great emphasis on regulations applicable to ML models in the US on different levels i.e. state and federal. It is important to understand them as the US government holds significant regulatory power after the McCarran-Ferguson Act of 1945. Federal regulators are taking a more hands-on approach when it comes to machine learning in insurance.

State Regulation

State departments of insurance (DOIs) are active in monitoring the process of machine learning implementation in insurance. They hold around 50 different interpretations with regards to different aspects of ML integration. However, their focus is on two primary aspects. Firstly, they enforce that fair dealing is practiced by the insurers with their customers. Secondly, they are concerned whether insurance companies have enough capital reserves to pay off claims at any point. Moreover, they also try to streamline the ML models of underwriting so that the results are not exploitive.

Despite the regulatory differences across different states, the general consensus is that insurance companies must explain the rationale of their underwriting decision. There are different restrictions at play in different states when it comes to underwriting decisions. For instance, some states expect the ML model to be made open to the public for any further analysis. In Hawaii, the use of customer's credit data is prohibited in auto insurance. Companies that have simple models find the explanation process to be relatively easy but as the model gets complex, explaining it becomes more challenging. That is why it is recommended to match the investment in compliance with the degree of complexity of the ML model.

Federal Regulation

The federal government plays various key roles in a supporting capacity for the insurance companies. It provides direct insurance to high risks markets such as flood and terrorism insurance which are simply not financially viable for regular insurance companies. Moreover, they also cover institutions which are viewed as too massive to fail such as banks. These companies are known as systemically important financial institutions or SIFIs. The health of SIFIs is monitored by Federal Insurance Office and Financial Stability Oversight Council which was established by the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010. Currently, Prudential is the only active insurance for SIFIs.

Moreover, through the Dodd-Frank's Non-admitted and Reinsurance Reform Act, the federal government exempts certain groups and provides support to reinsurance markets. It also prohibits unfair, deceptive, or abusive acts or practices (UDAAP).

The customer information as also protected by federal law in the insurance domain. Compliance of standards such as the Federal Credit Reporting Act (FCRA), the Equal Credit Opportunity Act (ECOA), and the Gramm-Leach-Bliley Act (GLBA) is necessary. For FCRA, insurance companies must have allowable reasoning to use customer information while medical information cannot be used without customer's permission. In case of denials or terminations of policies, customers are entitled to receive explanation for such actions.

The U.S. Federal Trade Commission and the Consumer Financial Protection Bureau are responsible for the enforcement of FCRA and ECOA. The purpose of ECOA is protect customers from discrimination on the basis of religion, sex, creed, age and many other personal parameters. ECOA also requires insurance companies to notify customers about any adverse action within 30 days and clear reasoning for the action is necessary as well.

GLBA further regulates private customer information. It lays the standards on how customer data pertaining to their unique identity must be kept secure. Guidelines are given to companies about handling and storage of customer information. It also prevents transfer of sensitive information to third parties.

Existing ML Guidance

The US federal government has various departments which are involved in the ML process used for insurance underwriting. Recently, the Federal Deposit Insurance Corporation and the Federal Reserve has followed in the footsteps of the Office of the Comptroller of the Currency (OCC) which has supervised Model Risk Management (MRM) since 2011.

These departments have broken the model down in 3 components which includes input, processing and output of results. Furthermore, they have also defined the risk of the model in 2 ways. One involves errors in the application of theoretical and mathematical concepts while the other focuses on use of obsolete models for solving new challenges.

To minimize the aforementioned risks, the Federal Reserve and OCC has developed process of validation which comprises of 3 parts. It starts with checking the model for any loopholes to ensure that the assumptions are reasonable and the limitations are identified. This information is used to create the necessary documentation. The second part is concerned with monitoring the inputs of the model and testing them against alternative approaches. Thirdly, the output of the model is analyzed and the results are compared with similar existing models.

MRM consists of detailed documentation regarding different topics. Definitions of mathematical theories and comparisons of the said theories with their alternatives is done. It also explains where proxy data can be used so that the errors do not exceed the set limits. The details in MRM also emphasizes that the model needs to be regularly monitored so that the necessary adjustments can be made.

Through the Federal Reserve's Consumer Compliance Outlook (2017), the Federal Reserve System explains how ML can comply with the existing regulations. The use of customer credit information for underwriting must be in line with ECOA and the Fair Housing Act (in cases of home insurances). Collectively, both regulations intend to protect the public from nearly all kinds of discrimination.

The Fed's Consumer Compliance Outlook has provisions to ensure that there is no bias within the ML model. These provisions focus on protecting the customers from disparate treatment and impact. The disparate treatment occurs when a customer is treated differently on the basis of a protected right. While the disparate impact occurs when the insurance company, intentionally or unintentionally, creates an impact on the customer that underwhelms a protected right. The guidance from 2011 was revised in 2017. Companies are expected to be in compliance with ECOA and FHA while avoiding UDAAP related risks. Companies, therefore, must strive to develop bias free ML models which are monitored regularly for any violations in the regulations.

International Regulation

Similar to US federal regulations, international insurers exercise regulatory action as well. In the EU, there is the European Insurance and Occupational Pensions Authority (EIOPA) that monitors the liquidity of companies and protects the rights of consumers. EIOPA defines its practices in both the Solvency II Directive and internal model directive. Perhaps the most important international regulation is the General Data Protection Regulation (GDPR).

GDPR places certain restrictions on insurance companies when it comes to ML. Since companies utilize growing amounts of data to make decisions, the responsibility of explaining those decisions also falls on them. It is considered to be imperative that a ML model must comply with the GDPR by having transparency of decision making process. Therefore, regular documentation and monitoring must be done to ensure that interests of the customers are not harmed.

GDPR has broken the explanatory requirement for insurance companies into sub-parts. First, companies are required to protect and correctly utilize customer's personal information like their credit history. Second, companies must have records which may be subjected to audits for checking compliance with GDPR. Third, companies must try to eliminate bias and prejudice in their decision making.

The article 4 of GDPR pertains to customer profiling as, in ML models, data is processed by an automated system. It defines profiling as an automated form of processing and evaluating customer's personal data in order to reach decisions. Regardless of its origin, a company must comply with GDPR profiling requirements if they cater to EU citizens. Article 21 and 22 of GPDR places strict regulations on profiling activities if they have a significant legal impact on the customer. Authorization from the customer's country of residence is needed for profiling. Moreover, the company has to enter in a contract with the customer after obtaining legal consent and giving the choice to opt out.

Compared to US regulations, GDPR places lesser restrictions on companies before taking any adverse action on their customers. Articles 13 to 15 give customers the right to view the logic and rationale behind the automated decisions in order to fully understand their expected ramifications. Therefore, companies who maintain a reasonable amount of transparency in their ML model should face little to no issues. On the other hand, the GDPR is similar to the US regulations with regards to protecting the customer base from any discrimination. It prevents the company from using personal data such as racial origin, religious beliefs and sexual orientation at the time of inputting information in the ML model.

The Black Box

While the use of ML in insurance is being explored, not many are able to implement it successfully on the underwriting process. It has been already highlighted that the laws and regulations are considered as an impediment for ML deployment. Similarly, the complexity of the model poses a challenge as well. It is observed that understandability of the model has an inverse relation to its predictive power.

The ML model outperforms the standard logistic regression model used by the insurance industry. The logistic regression model uses fewer variables, normally less than 50, when making decisions. On the other hand, the ML models are not only capable of using thousands of variables but they can develop cross-links between them. The quality of the results improves with the increase in the number of variables which ultimately leads to better analytics and decision making. Moreover, the risk assignment to different variables is more nuanced compared to the generic and idealistic risk assessment of the logistic regression model.

In order to practically test the effectiveness of ML, some vendors have started to offer ML in a sandbox to insurance companies. This allows data scientists of different organizations to gain firsthand experience of applying their business data in an internalized environment. In most cases, the results are quite positive but companies often abandon the use of these models in real life when they find out that they do not comply with some of the regulations. The details provided by the “black box” put off many companies from adopting ML models as they feel it might be too much of a hassle to reconcile them with the relevant laws and regulations.

Change Management

It is a common notion that complications in the adoption of machine learning normally originate from constraints in mathematical models and compliance. However, it can be argued that organizational rigidity is the greatest challenge when it comes to introduction of machine learning. At times, the people present within an organization are simply not willing to go through the transition process. The change management can be broken down into 3 parts: talent, program management and organizational alignment.

Talent

There is a large disparity between the demand and supply of ML experts in the current job market. As a result, it is very hard to find the right employees for handling your machine learning needs. If the right employee is found, they may have high salary demands and can cause issues when it comes to retention.

ML professionals need a vibrant environment with a strong ML infrastructure. Not meeting their expectations can lead to their departure as such employees prefer to work in an ML oriented workplace in order to boost their personal and career growth.

IT Program Management

A powerful ML model can only work and thrive if it is backed by a sound IT infrastructure comprising of reliable hardware/software and a competent workforce. In terms of IT program management, one may come across a host of challenges which can be broadly divided into technical and non-technical problems.

Most of the technical issues originate from ensuring that you have the right specifications that meet your ML needs. It is important to decide which open – source programming language (like Python and R) would work best with your ML model. Then comes the choice of how the data would be managed and what kind of computing, cloud or hardware-based, would be used to implement the ML model. Again, the costs and features of both possibilities need to be understood first. The computing for a ML model can be divided into 3 phases: training, scoring and monitoring. All three have different computing needs which have to be met. Moreover, latency and internet speed are also important points of consideration when it comes to technical challenges.

Moving to non-technical challenges, hiring the right IT technical support personnel and defining the parameters of the solution are problematic. A timeline needs to be set for different modules of the implementation. Moreover, delegating tasks to the right person is important for ensuring success. IT personnel are responsible for compiling the data sources, both internal and external, which will be used by the ML model. They also need coordinate with different departments to ensure that the output of the model is understandable to the decision makers. Normally, the owners and higher management are involved in the process so that they can make optimal decisions.

Organizational Alignment

Aligning the organizational hierarchy is critical for a smooth transition to ML models. Liaison and coordination is needed across all relevant departments such as IT, analytics, data science etc. Challenges can vary depending on the structure followed by the company but there are some general issues.

Unnecessary friction can develop among different departments if the implementation of the model is not in-sync which causes some parties to lag behind in the process. It can lead to the blame game among departments in cases of delays. That is why it is important to have a robust timeline which has the necessary slag time to allow for delays along with proper benchmarking of milestones. It is common to find that a certain department may feel threatened or incompetent if their work is outsourced. Clear reasoning should be communicated in such circumstances to avoid decline in workforce morale.

Alignment issues can also be resolved with the aid of external help. Experts which specialize in implementing ML models can be brought on board as they are familiar with the common challenges, and their relevant solutions, during the transition. We will further explore the benefits of external help in the last section.

Implementing Machine Learning

Before starting with the implementation, it is important to assess the current competencies of the staff with regards to handling machine learning models. In case of any gaps, third-party machine learning experts can be hired for training the staff. In some cases, a customized machine learning model is needed so, in that event, machine learning professionals may be called upon once again. Furthermore, if the relevant model already exists, machine learning expertise can help in getting the process up and running. Tasks such as explaining the “black box” in order to meet the regulatory standards can be handled expertly by them.

The next step would be to map the ML model to your regulatory and enterprise model risk management (MRM) requirements. The development of a proper MRM requires extensive documentation. It defines the methodology used to develop the model, usage of the data and protection of customer information from any misuse or exploitation. To ensure that the criteria of your existing validation model is met, in-depth review of the documentation is necessary. Since the reports for MRM can be quite lengthy (comprising of nearly a thousand pages or so), it is wise to explore the automation option for generation of these reports. This not only helps the insurance regulators maintain transparency within the system but it also allows to focus valuable resources on the bottom line and other important tasks like innovation and improving customer experience.

Lastly, it is important to setup a monitoring and alerts protocol which prevents deviation from the ML model's purpose. It is important that the ML model maintains its bearings and remains on course to meet the set deliverables. It is also imperative that activities pertaining to machine learning do not harm the interests of the customer base. That is why it is recommended that the models are designed with these points in mind and, once in production, they are regularly monitored. It is a good idea to conduct regular automated tests of the model. These tests can compare the economic performance of the model with its contemporaries. On the other hand, they can also send out alerts if the customer base response changes negatively so that corrective action can be taken.

Conclusion

Machine learning can revolutionize the insurance business by improving the utilization of existing data within the industry. The implementation of cutting edge ML models can increase profits and lower costs. The underwriting process has drawn great benefit from the introduction of machine learning. Customers with limited value are removed from consideration while the focus shifts towards higher value customers who are discovered through the usage of machine learning.

It is important to navigate and circumvent the challenges linked with use of machine learning in the insurance industry. There are a host of regulations and laws that require your compliance. The idea behind them is to bring the necessary transparency that can give an internal perspective of the "black box". Uncovering the box explains the processes and methods used to reach decisions using the technology. This transparency also becomes useful when overcoming the challenge of change management. The higher ups will only be willing to adopt ML if they understand how it was used to reach a particular conclusion. Once they are convinced, it becomes a trickle-down effect where the rest of the company is more open to accepting ML.

At the end of the end of the day, the difference between getting a top of the line ML system accepted or rejected by the organization is its ease of understandability. With Axilient, you get simplified yet modern ML solutions which have the necessary features to alleviate the concerns of all stakeholders including management and regulatory authorities. Axilient has helped many insurance businesses in changing their fortunes through modernization of the underwriting process.