6 – Data Analytics

# 1 – Big Data and Technology – The Future of Insurance

Objective: Describe the evolution of big data and technology and its impact on the property casualty insurance business.

Technology has transformed our lives in how we communicate, learn, and do business. It also continues to evolve at a rapid pace. Example, self-driving cars are now a reality. Technology has already changed the insurance industry, and these changes have only just begun. The convergence of big data and technology has started a transformation of the property-casualty insurance business.

Insurance would not be feasible without data. The ability of insurers to provide coverage for a variety of risk exposures is based on the law of large numbers. Traditionally, the data insurers use has come from loss histories. By analyzing large numbers of claims, insurers could reasonably predict the probably future of claims.

However, as technology evolves, the sources and amount of data available to insurers are increasing rapidly. To effectively use this data, insurers must continue to develop new methods of obtaining, processing, and analyzing it.

## Strategic Opportunities from Big Data and Technology

Why would an insurance professional who is not an actuary or a data scientist want to learn about analytics? The most comprehensive reason is that **big data and technology are central to the future of the insurance industry. The evolution of this area has already disrupted the traditional ways in which insurers market, underwrite, and analyze their products.** This disruption will only gain momentum with such changes on the horizon as self-driving cares, sensors in products ranging from bridges to jean, and computers that teach themselves without human intervention.

**Another more practical reason is that insurance professionals, such as those in underwriting and claims, must be able to communicate with the data scientist in their organizations**. Those who are engaged in the typical work or an insurer may serve on multifunctional teams to assist in the design of approaches to data analytics. **Additionally, professionals in traditional insurance functions will need to understand how to use the techniques developed by data scientist to make decisions. Also, as the application of data analytics to insurance continues to evolve, professionals in areas such as underwriting and claims will be valuable in evaluating the results of the new techniques**.

Data-driven decision making has been proven to produce better business results than other types of decision-making. Example, the credit industry was using data to select clients and determine credit lines. The data was obtained from loss histories. A small bank with little or no loss histories decided to accept credit randomly and suffered significant losses. The bank acquired enabled tis data scientists to develop predictive models that identified the characteristics of the most profitable clients. The bank then use the results to develop a strategy regarding which credit products to offer to which clients as what level of credit and interest rate. That bank became Capital One.

## Big Data

**Although insurers have always gathered and analyzed large amounts of data to make decisions, the amount of data available has increased exponentially and is therefore referred to as big data. Two types of data are available to insurers. The first is their own internal data, and the second is data from external sources**.

Why would insurers’ own internal data now be considered advancements that require new methods of analysis? The answer is *based on techniques such as* ***data mining (analysis of large amount of data to find new relationships and patterns that will assist in developing business solutions****) and text mining. Example of mining: using new technique to identify previously unknown factors that are common to an insurers’ most profitable auto insurance customers. Other advanced techniques can enable innovative ways to market to those customers. Text mining can be used to analyze claims adjusters’ notes to identify fraud.*

**New external sources of big data are constantly evolving. Example; Telematics. The data provided by telematics is analyzed to determine the driving patterns that lead to accidents as well as reflect safe driving. This information can result in real time change in premium if the auto owner agrees**. A safe driver could receive a discount, while an unsafe driver may receive a premium increase and/or be required to attend remedial driving program. The information obtained from telematics can also be applied to an insurer’s online application and selection process.

## Technology

Technology is constantly providing innovative products that are changing risk management and insurance. Similar to telematics devises, sensors can provide a stream of data to insurers regarding risk. Heat sensors can provide information about fire risk to both the customer and the insurer. Sensors on machinery can indicate a malfunction or the removal of a safety device. Sensors are now being included in clothing, which are referred to as wearables. Wearable in uniforms can provide valuable ergonomic information to improve safety.

*An emerging area of technology is the Internet of Things (IoT). Although this technology is still being explored for its potential in insurance, it is likely to evolve rapidly. The IoT is similar to telematics, home sensors, and wearable technology, but its potential is more extensive. The potential involves machine to machine communication and machine learning*. **Machine learning is artificial intelligence in which computers continually teach themselves to make better decision based on the previous results and new data**. The IoT may eventually be able to identity the next asbestos-type problem before it occurs by analyzing data from wearables, sensors, and text mining. Example, nanotechnology is a relatively new type of process, and its eventual effects are still unknow. The IoT could help insurers monitor risks associated with this process.

*Drones are a new technological source of information, and they can produce data for underwriting or claims management. Instead of sending adjusters to a disaster scene, insurers are now testing the use of drones to provide detailed data about property damage. Drone, along with other technological tools such as cell phone and the internet, can assist in gathering risk data for underwriting property coverage in wildfire-prone areas*. Results from modeling data already obtained have proved the new, valuable insight that a well-constructed, well-maintained property in a high-risk area can be a better risk than a poorly maintained one in a moderate risk area.

Social media is a rich source of data. Google and Facebook target advertisements to people who are likely to be interested in a particular product based on their profiles on social media sites. Google, Facebook and Apple are also developing artificial intelligence (AI) – computer processing or output that simulates human reasoning or knowledge – to find answers to rare search inquiries. All these organizations are researching techniques that will provide an analyze information from voice recordings, photos, and videos.

AI has significant potential for insurers. Claims adjuster often take recorded statements from claimants. AI techniques could be applied to recognize voice patterns that may indicate the possibility of fraud. There could also be applications related to photos of accidents or natural catastrophes. Whether and to what extent insurers can use social media, such as videos posted on Facebook is subject to regulatory and privacy concerns,

## Data Science

**Data science is a new field that arose from the need to link big data and technology in ways that provide useful knowledge. As discussed earlier, big data by definition, is an amount of data too large to be analyzed by traditional methods. Data Scientists have interdisciplinary skills in mathematics, statistics, computer programming, and sometimes engineering that allow them to perform data analytics on big data**.

Data scientist use various techniques to organize and analyze data. Machine learning is on example. Traditional techniques, such as probability, are often combined with newer techniques to obtain the most relevant knowledge from the data. **A major concept in data scient is that the data organization and analysis are automated rather than performed by an individual**.

**However, human evaluation of automated data analysis is critical**. First, analysis performed by computers is not always accurate. Example, Weather forecasting. Different models interpret meteorological data differently and produce varying forecasts. Experienced meteorologist must evaluate all of the model analyses and use their professional judgement to make a decision. Second, the automated analysis may be correct but irrelevant to a given business problem. Example, automated analysis of claims indicating that teenaged male drivers are more likely to have accident would not be relevant to a deeper understanding of auto accident risks because this information is already known. Third, as technology is evolving, so are the physical, political, economic, and business environments. Unless the automated method can take those factors into account, the results of the automated process may not be helpful if the environment suddenly changes.

For data science to be useful to an insurer or to specific functional areas, such as underwriting or claims, is usually important to define the business problem to be solved, such as improving claims or underwriting results. In the future, data analytics may be used to forecast unforeseen, or “black swan”, events.

## Strategic Applications to Insurance

Insurers and risk managers should determine which investments in big data and technology will provide the best fit for their business. Example: An insurer that provides only personal lines of coverage may not be interested in wearables. A company that operates a fleet of vehicles may be more interested in telematics than in building sensors. Insurers and risk managers also need to decide on the computers and software to process the increasing amounts of data.

**After the data has been obtained and processed, insurance analytics must be applied to make data-driven decisions and develop strategy. This involves not only the big data, technology and data science techniques but also people who evaluate the results of the analytics**.

Insurance professionals in areas such as underwriting, claims, and risk management should have input into the types of data that they know would be helpful to their decision making, such as fraud indicators. These professionals should also have input into how well the analytics are working to produce relevant and practical information or to improve efficiency.

At more senior levels of an insurer, the information obtained from data analytics can guide decisions such as which line of coverage management wants to increase or decrease, countries where it would like to expand or reduce operations.

## Evolution of Big Data

Big data and technology can be regarded in stages of evolution. The first Bit Data 1.0 which is the stage at which many organizations are now. The next stage is Big Data 2.0. Some organizations are already at this stage. There will certainly be future stages as technology continues to evolve.

### Big Data 1.0

**In Big Data 1.0, organizations began using the internet to conduct business and compile data about their customers. Insurers developed online insurance applications. They also use data from applications, as well as claims history, to improve underwriting efficiency and to provide customer information for product development and marketing**.

Other industries were performing similar functions in Big Data 1.0. For example, retail firms, such as E\*Trade, could use the internet to provide new products to customers. Computers could place stock and bond trades much more efficiently than previous methods.

### Big Data 2.0

**Because of rapid advances in data science, we have reached the stage of Big Data 2.0. This stage allows organizations to obtain and aggregate vast amounts of data very quickly and extract useful knowledge from it.**

Example: Investment firms have computers that can scan the internet instantly for new and information about products, prices, economic and geopolitical developments, and consumer trends. This data is then provided to computer trading algorithms that conduct automated trading of stocks and bonds at high speeds.

**Big Data 2.0 also allows organizations to process and analyze data from sources such as vehicles, homes, and wearable technology**. Sensors can be built into products, and data science provides methods to process an analyze the data that comes from those sensors. Big Data 2.0 is evolving quickly. Some insurers are already actively involved in this phase, while others are lagging. It is not an exaggeration to say that insurers that can adapt quickly and effectively to the convergence of big data and technology will have significant competitive advantages.

# 2 – Big data Characteristics and Sources

Objective: Categorize the various characteristics and sources of big data available for insurance and risk management applications.

Insurers’ traditional data is organized into databases with defined fields. Insurance professionals can produce reports that show results from this data. Risk managers can usually access their insurer’s data to produce reports about claims they are handling. Big data has introduced different types of sources of data than those traditionally used.

It is important for risk management and insurance professionals to understand the different types of big data. The varieties, volume, an sources of data are rapidly increasing. Example: Risk managers can obtain new types of data on safety from sources such as sensors on employees. Underwriters can obtain new data from drones and social media. Claims adjuster can better identify fraud by identifying patterns in internal and external data. Risk management and insurance professionals increasingly work with data scientist to determine the types of data that are useful to make new business decisions.

## Data Characteristics

The term “bid data” implies large quantities of data. Although this is true, big data is also different from traditional data in other respects.

There are characteristics that differentiate big data from traditional data

* Volume – there is an enormous amount of data that is now available, and the amount continues to increase.
* Variety – traditionally insurers used structured data (predetermined defined fields). Big data also includes structure data in larger volume than traditional data. However, because big data comes from multiple sources, much of it is unstructured (not organized in predetermined defined fields).
* **Velocity – this is the constantly increasing speed at which data arrives at an insurer compared to traditional data. The velocity of big data also includes the growing rate of change in the types of data compared to traditional data**.
* Veracity – this refers to the completeness and accuracy of data. Unstructured big data is more likely to have less veracity.
* Value – value is derived from the results of data analysis to help insurers make better decisions. Big data has great potential to add value, but it must be obtained and analyzed with techniques that provide meaningful results. This is the goal of data science.

## Internal and External Data

Data science allows insurers access to increasingly larger and more varied data referred to as big data. Some of this data is received directly by the insurer or risk managers’ organization. Other data is obtained from outside sources.

### Internal Data

Insurers have always relied on data, and they possess large quantities of it. Certain types of data, such as risk factors, losses, premiums, rating factors, rates, and customer information, have been traditionally used to make business decisions. Much of this data is also reported to state rating bureaus. Similarly, risk managers also have relied on data about losses and premium, in addition to safety statistics that are reported to OSHA.

However, risk managers and insurers have vast quantities of internal data that they have not used because of its volume and/or lack of techniques to access it. Both traditional and nontraditional internal data that could be useful if data science techniques are applied to it. Example: analysis of certain medical codes could provide useful information to an insurer’s claim department or risk manager. Nontraditional data, such as recorded statements and photos, could also provide additional information through data science.

Data science also provides techniques to use nontraditional internal data. Example: voice analysis can be applied to recorded statements of claimants to identify vocal characteristics associated with fraud, such as gaps in the claimant’s version of the accident or a defensive tone. Artificial intelligence can analyze photos to find information that may e missed by human analysis, such as a potential environmental hazards around a property or, through facial recognition, the identity of a witness to an accident.

### External Data

**There is sometimes a blurring of the boundary between internal and external data. For example, data from telematics is obtained from a devise that is installed on a customer’s vehicle. The insurer owns the device, but the customer owns the vehicle. Because of the customer’s ownership of the vehicle and nature of the data provided – the customers’ personal driving habits – telematics should probably be categorized as external data**.

Another source of data with a blurred boundary is a wearable sensor used by an employee while working. The employer provides the sensor, but it is placed on the employee’s person and transmits information about the individual. This information could be considered internal. However, it could also be considered external because it is obtain from and about the individual.

Other sources of external data include sensors in customer’s homes or commercial buildings, photos of a catastrophe obtained by drones, and the internes- especially social media. For insurers to access information from sensors on customers’ property, customers must first grant permission. Regulators have already approved the use of drones by insurers to inspect catastrophe scenes. However, there may be other regulations governing what information can be used and how it can be used.

**Privacy issues are a legal and regulatory concern in obtaining mass information from public sources, including social media. This is particularly true with data science because it is used to discover and analyze previously unknown data relationships. For example, an insurance company might access individual items of data on a claimant from both internal and external sources. While the data from each source by itself might not violate the claimant’s privacy, combining all the data might leas to conclusions that harm the claimant’s privacy in a way that he or she did not anticipate. As a result, privacy regulations have been passed by various states and countries.**

Other sources of external data include statistical plan data, other aggregated insurance industry data, competitors’ rate filings, and third-party data. The most commonly used types of third-party data are economic data (data regarding interest rates, asset prices, the consumer price index, and other information about global, national or regional economy) and geodemographic data (data regarding classifications of a population).

## Structured and Unstructured Data

**Data can be either structured or unstructured. The traditional data that insurers use for underwriting and claims management is structured data. This data is organized into databases that are often linked to each other. Telematics provide an example of structured external data**. There are predetermined fields in a data base for the driving information that insurers receive from telematics, such as distance traveled, braking, and left turns.

**Unstructured data is not organized. Adjuster’s notes are an example of internal unstructured data. Although the notes are contained in individual claims files identified by numbers, there is no database that categorizes the data contained in the notes** and provides results. Unstructured external data includes information from the internet, such as social media.

Data science is used for both structured an unstructured data, but it is especially useful for unstructured data. Data scientist use various techniques to gather, categorize, and analyzed unstructured data. This process may or may not produce useful results.

**Big Data Categories and Examples**

|  |  |  |
| --- | --- | --- |
|  | **Structured** | **Unstructured** |
| External | Telematics  Financial Data  Labor Statistics | **Social Media**  News Reports  Internet Videos |
| **Internal** | Policy Information  Claims History  Customer Data | Adjuster Notes  Customer Voice Records  Surveillance Videos |

# 3 – Data Mining

Objective: Describe Data Mining and the Data Mining process

Big data can be very useful to insurers and other organizations. However, it is only useful if it can be properly processed and analyzed.

Data Mining is the application of various techniques to big data. Data mining and the data mining process allow organizations to analyze large amount of internal and external data from various sources.

## Data Mining Techniques

Data mining is closely related to the fields of statistics, machine learning, and database management.

**Although data analysts use different approaches, depending on the nature of the data and the business problem, these are the basic techniques of data mining:**

* **Classification – assigning numbers of a dataset into categories based on known characteristics**. (similar to class codes of insureds, or claim class codes)
* **Regression analysis – a statistical technique that predicts a numerical value given characteristics of each member of a dataset.** (example to learn the % of premium increase that would lessen the customers’ likelihood of renewing an auto policy)
* **Association rule learning – Examining data to discover new and interesting relationships**. From these relationships, algorithms are used to develop rules to apply to new data. An insurer can explore data to find relationships among its products purchased.
* **Cluster analysis – using statistical methods, a computer program explores data to find groups with common and previously unknown characteristics**. The results may or may not provide useful information.

**Classification and regression analysis are data mining techniques that are applied when an insurer knows what information it wants to predict**. The insurer also knows the specific characteristics of the data it plans to analyze to find the answers to these questions.

**Association rule learning and cluster analysis are used to explore data to make discoveries. An insurer might apply cluster analysis to discover customer needs that could lead to new products. Unlike classification and regression analysis, there are no known characteristics of the data beforehand. The purpose of association rule learning and cluster analysis is to discover relationships and patterns in the data and then determine whether that information is useful.**

## Developing a Predictive Model

**A data mining techniques is applied to a set of data. The result of this is a model that can be applied to new data.** In the example about determining whether a claim might be fraudulent, a classification technique can be used to categorize each claim. After the results are evaluated and validated, the model is used on new claims data to indicate if the claim needs to be sent for further analysis**. As the model is applied over time, machine learning refines the model by analyzing the newly added data to revise algorithms to better predict results.**

## The Data Mining Process

In 1999, a consortium of individuals who worked in the emerging field of data mining at Daimler Chrysler and SPSS received funding from the European Union to develop a data mining standard**. The result was the Cross Industry Standard Process for Data Mining (CRISP-CM).**

A fundamental feature of this standard is the circle that surrounds the process diagram. This circle indicates that data mining is a continuous process that involves continually evaluating and refining the model.

**The first step in the process is to understand what a business wants to achieve by applying data mining to one or more sets of data**. Does the insurer wish to retain more customers or better identify fraud? **This step is especially important because without a business context, modeling and analyzing data is unlikely to be effective.**

**The next step is to understand the types of data that are being used**. Is the insurer using only its own claims data, or is it also using data from statistical rating agency? Is the only data being used from sensors and its own claims data, or is the organization partnering with an insurer to obtain and analyze data?

**After the data is selected and its source(s) understood, it must be prepared**. This involves cleaning the data as much as possible to eliminate missing or inaccurate data. The purpose of the data mining determines to what extent the data must be cleaned. Example: if it is being used to price insurance the data should be high quality. If it is exploring social media data to look for opportunities for new products, the data may be usable even if it is lower quality.

**After these steps, data mining techniques are applied to develop a model to analyze the data. The results are evaluated to determine whether the are reasonable and meet the business purpose**. Even if they do not meet the purpose, the data mining may still be of use. The data might be better prepared or the model adjusted to produce the desired result. Even if the model produces useful information**, machine learning will continue to refine that model to produce increasingly better and more accurate information after the model is deployed to analyze new data**.

# 4 – Data – Driven Decision Making

Objective: Explain how data-driven decision making applies to risk management and insurance

Risk managers and insurers have traditionally made decision based on data. Insurers use their own and industry loss data to price insurance products. Risk managers use accident data to reduce losses. However, data analytics can improve the types of data, methods of analysis, and results by applying data-driven decision making. – An organizational process to gather and analyze relevant and verifiable data and then evaluate the results to guide business strategies.

## Data Science and Data-Driven Decision Making

Data science can help insurers and risk managers improve their business results through data-driven decision making. These are several ways to achieve improvements:

* Automating decision making for improved accuracy and efficiency – providing online quotes
* Organizing large volumes of new data – organized data according to multiple characteristics, such as telematics
* Discovering new relationships in data – workers who never had a WC injury
* Exploring new sources of data – using text mining to develop an automated system to predict claim severity

Across all types of businesses, perceived improvement in companies using big data has increased by 26% and is expected to increase by 41% in the next few years. Currently, insurers process only approximately 1—15% of the available structured data.

**The data about using bid data and data science for decision making is compelling. Insurers and risk managers have access to a wealth of existing data they are not tapping. Additionally, there is a rapidly growing quantity of external data from various sources, such as the internet at large and the Internet of Things IoT**.

**Insurers, as well as other organizations, acquired data engineering and processing technology that can manage big data, which is too large for most conventional systems. Models can then be developed to gather an analyze data in the context of an insurer’s areas of interest, with results ultimately provided to the data analytics team or the manger who requested the data analysis. The appropriate person(s) can make data-driven decisions accordingly**.

*Data-driven decision making can be applied across an insurer’s or risk manager’s enterprise to solve a variety of business problems, achieve greater efficiency, and provide a competitive advantage.* There are two basic approaches to data-driven decision making: Descriptive and Predictive.

## The Descriptive Approach

**The descriptive approach is applied when an insurer or risk manager has a specific problem. Data science is intended to be used to provide tat that will help solve the problem. Insurer or risk managers do not continue to use data-driven decisions beyond the specific problem**.

Example, the insurer wants to obtain data regarding accident rates for insureds who were accepted for coverage during the past year and who qualified based on 3 years without an accident.

The insurer intends to use the results of this analysis to determine if its new 3-year guideline is producing acceptable results. Using data-driven decision making, executive will decide whether to continue the three-year guideline or return to the five-year guideline. This is a one-time problem for data-driven decision making, and the analysis will not be repeated.

## The Predictive Approach

**A predictive approach to data analytics involves providing a method that can be used repeatedly to provide information for data-driven decision making by humans, computers or both**.

Example, automated underwriting for personal auto insurance is a predictive approach that is used each time a person applies for insurance.

**In addition, predictive analytics can be used to assign claims to an adjuster.** However, the claims manager may have authority to reassign a claim that the computer assigns as a minor claim if the claim turns out to be severe.

**Risk Management and Insurance Data Analytics Decision-Making Model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
| **Define Risk Management of Insurance Problem** | Data | Analysis and Modeling | Insights | Actionable Decisions |
|  | Data Selection  Data Quality  Data Preparation  Data Understanding | Big data analysis techniques  Descriptive and predictive models | Trends  Relationships  Behaviors  Events | Risk pricing  Risk Selection  Claim Fraud Det  Claim assignment  Product develop  Risk identification  Risk assessment  Loss control |