

E-learning modules

DATA CULTURE TRAINING PROGRAM

EN



E-learning: Introduction to data culture

20 min – Beginner level

Abstract

This module introduces the functioning of an insurance company which goal is to transform a situation of uncertainty into a situation of near certainty. It helps to understand why mutualisation and diversification are key to it functioning and which risks are created by the insurance business. We also introduce the importance of data all along the insurance value chain.

Learning goals

- Explain the inversion of production cycle of an insurance companies
- Present the 3 main principles of functioning of an insurance company: mutualisation, diversification and time value of money
- Provide examples of use of data all along the insurance value chain

Agenda

1. Functioning of an insurance company
 - Inversion of production cycle
 - 3 natural protections: mutualisation, diversification and time value of money
 - Evolution of the world
2. Insurance value chain
 - Data all along the value chain
 - Examples of use
3. Structure of the training
 - Data culture
 - Data practices
 - Innovation

E-learning: Main principles about data

45min – Beginner level

Abstract

In this module we briefly present what is data science and its origins. Then we describe different applications and interest within industries focusing on the insurance area. We introduce the current dynamic ecosystem of data science communities focusing on actors but also on tools, typical events and careers. Finally some limits and risks will be highlighted.

Learning goals

- Paint the scene of the current data science ecosystem
- Understand origins and why data science is such a crucial topic today in insurance
- Have a global overview on main elements of the ecosystem and understand interactions
- Keep in mind limits

Agenda

1. What is data science
 - Main principles
 - Origin of data science
 - Main methodologies
2. Awareness of different industries
 - From a global point of view
 - Zoom on financial sector
 - Main application and interest in Insurance
3. A large connected ecosystem
 - Community animated by different kind of actors: fintech, startup, professional community, etc.
 - Evolving thanks to a large range of tool
 - Praised through many networks: blog, hackathon, meetup, professional colloquium
 - Motivating and challenging career perspectives
4. Behind the scene: limits to keep in mind
 - Ethic: personal or sensitive data are sometimes used to create project.
 - Legal
 - Security
 - Market perspectives
 - ROI and Costs

E-learning: Data preparation and data quality

40min – Beginner level

Abstract/Context

In this module we explain how data is created, gathered and manipulated to be optimally used in companies. We also propose different best practices to address as efficiently as possible preparation of data with a view to creating data science projects.

Learning objectives

- Examine all sources of data, its structuration and its different shapes
- Know how data is manipulated and organized within a data driven company
- Highlight some best practices about data preparation and manipulation

Agenda

1. Source of data
 - Many sources:
 - Internal data : teams, clients, provider, website
 - External data : social network, open data, IoT, scraping
 - Structured versus non structured
 - Different types of data: Numerical, Textual, Audio, Image and video
 - Data creation through the value chain of an insurance company
2. Organization and management of Data
 - General process
 - Creation of data
 - Gathering of data
 - Transformation and controls
 - Final use
 - Re-use
 - Focus on Data warehouse and data lakes
 - Data quality principles
3. Main quality checks and actions
 - Missing values
 - Coherence and accuracy check and abnormal values
 - Statistics, discretization, correlation
 - Dimension reduction
 - Feature engineering

E-learning: Data governance within Solvency 2

40min – Beginner level

Abstract/Context

In this module we introduce the main elements of the Solvency 2 framework and its impact on data management. We explain the typical process and the key roles around data governance.

Learning objectives

- Explain the main concepts of the Solvency 2 framework
- Present the requirements introduced by Solvency regarding data management, data quality and data governance
- Present the process and key roles around data governance and the related documentation that should be in place

Agenda

1. Introduction to the Solvency 2 framework
 - 3 pillar approach
 - Economic balance sheet: main elements
 - Solvency Capital Requirements
 - The 4 main functions in the system of governance
 - Reporting and disclosure
 - Main impact on data: data quality and data governance
2. Data governance process and key roles
 - Identification of the necessary data
 - Definition of the data architecture
 - Definition of data quality standards
 - Data governance structure
3. Documentation
 - Data quality policy
 - Data dictionary
 - Reporting

E-learning: Insurance regulation about Data

35min – Beginner level

Abstract/Context

In this module we present the General Data Protection Regulation (GDPR) that will enter into force in May 2018 and its impact on the treatment of data by (insurance) companies.

Learning objectives

- Introduce the main principles of GDPR
- Explain the obligations related to GDPR
- Create awareness among the collaborators of the company on the impact of this new regulation
- Present some practical applications in the context of an insurance company

Agenda

1. Context
 - Right to privacy and right to the protection of personal data
 - Technological evolutions
2. Scope of application, definition and principles of GDPR
 - Who is concerned?
 - Different forms of consents
3. Practical organization and Obligations related to GDPR
 - Transparency
 - Obligations
 - Responsibility
 - Penalties
4. Additional topics
 - Outsourcing data treatment
 - Data transfer to foreign countries

E-learning: Data analytics and statistics

45min – Beginner level

Abstract/Context

In this module we present some basic elements of the analysis of quantitative data and statistics and how these concepts are at the basis of the insurance business. The main goal is not to go deep into the details but be sure that the basic concepts and vocabulary are clear to the participants.

Learning objectives

- Introduce the concept of mutualisation and explain why it is central in insurance
- Develop useful statistical concepts that are at the basis of data analytics
- Present practical applications in insurance where statistical analysis is key

Agenda

1. Mutualisation and Statistics: the basis of insurance
 - The concept of mutualisation in insurance: gathering risks to reduce uncertainty
 - A pragmatic introduction to the law of large numbers
2. Useful statistical concepts
 - Random variables
 - General concept
 - Examples: number of claims and claim amounts
 - Elements of Probability
 - Mean and variance
 - Statistical distribution
 - Discrete vs continuous distribution
 - Some examples: Normal distribution and Poisson distribution
3. Methodologies families
 - Inference
 - Regression
4. Practical applications in insurance
 - Premium calculation
 - Capital requirements
5. Limits
 - Model error
 - Parameter error

E-learning: Introduction to Machine learning

50min – Beginner level

Abstract/Context

In this module we introduce machine learning concept from historical aspects to theoretical and practical perspectives. We present shortly main characteristics of this kind of statistical approach and how it can help insurer to improve/optimize its business.

Learning objectives

- Understand the interest of machine learning compared to a simple program
- Provide a high level presentation of different machine learning technics
- Propose a generic best practice process to address machine learning issues
- Present different machine learning application in Insurance

Agenda

1. Introduction
 - Artificial Intelligence origins: introduction and historical
 - Classical program versus machine learning:
 - Definitions: Model is defined within 3 questions: how to represent it? How to evaluate it? How to optimize it?
2. Algorithms families
 - Introduction : 4 main families within machine learning area
 - Supervised ML
 - Non Supervised ML
 - Other methods (partially supervised and reinforcement) :
 - Comparisons
3. General process/methodologies
 - Identification of the problem
 - Data preparation and split
 - Model error and optimization
 - Visualization
 - Continuity
4. Example of use in Insurance:
 - Supervised ML regression : Non-life insurance pricing
 - Non-supervised ML regression : Life insurance policies clustering
 - Supervised ML classification: General condition understanding
 - Non-supervised ML classification : broker performance study

E-learning: Text mining

45min – Beginner level

Abstract/Context

In this module we define text analysis practice. More specifically, we introduce how text mining process could be set up and how we can apply it to an insurance context to unleash power of document (both internal and external).

Learning objectives

- Understand how to manipulate raw text
- Learn how to define an accurate text analysis process
- Provide a few examples about insurance applications related to text mining

Agenda

1. Definitions
 - Introduction of text mining and Natural Language Processing template
 - Interest of this type of methods
2. General process
 - Global process from taxonomy to final use
 - Focus on corpus treatment and extraction:
3. Advanced methodologies
 - Part of speech method (POS)
 - Optical character recognition and image recognition (OCR)
 - Sentiment analysis
 - Models (unsupervised machine learning for document management) and visualization (wordcloud, ggraph, etc.)
 - Limits: language, synonyms, ironic sentences, information storage, model complexity
4. Examples of use in Insurance
 - Insurance General conditions comparison
 - Claim management request classification for process optimization
 - Reinsurance treaties automated reported treatment

E-learning: Open data

35min – Beginner level

Abstract/Context

In this module we present why open data is such a hot topic today, how does it work and how insurers can take benefit from these kinds of initiatives.

Learning objectives

- Understand what are open data
- Know how to manipulate these kind of content
- Have in mind a list of open data projects and applications in the insurance sector

Agenda

1. Open source context and motivation
 - Open movement
 - Free software movement from GNU project to R, Python
 - Rise of Internet
 - Open data government initiatives
 - Pro and cons about open source
2. Open data
 - General introduction : data that anyone is free to use, reuse, and/or redistribute
 - Public versus private initiatives
 - General topics: Economy, Education, Energy, Environment, Health, Justice, Science, Transport
 - Type of data
 - General characteristics
 - Focus on roles
 - Focus on use and distribution
3. References
 - Belgium reference
 - European reference
 - Worldwide
4. Examples of application
 - Reminder before using open data
 - Example 1: Cadaster dataset for household insurance
 - Example 2: Medicine consumption for health insurance

E-learning: Scraping

35min – Beginner level

Abstract/Context

In this module we present web scraping techniques and how this method is relevant for targeting, extracting and structuring unstructured web information. We also present some limits about this practice and provide reasonable advices to exploit it without troubles. Finally, we describe insurance applications using scraping tools.

Learning objectives

- Present main characteristic of scraping techniques
- Teach how to set up a simple process without troubles
- Provide several examples of relevant applications within the insurance sector

Agenda

1. Concept introduction
 - Definitions
 - Distinguish between crawling, scraping and parsing
 - Interest of scraping
 - Many sources: website, API, etc.
2. Main practice and process
 - Define a website target or a range of websites
 - Create proxy and scraper algorithm
 - Download document, parsing and extraction
 - Drop structured info within storage
 - Data treatment and final use
3. Limits
 - Website limitation (legal mentions)
 - Copyrights and jurisprudence (LinkedIn, Ryanair)
 - Security (DDoS)
 - Consequences: html remap, IP blocking, etc.
4. Advices
5. Applications within Insurance sector
 - House market prices for household insurance pricing
 - Car rent and motor insurance price for car leasing offers
 - Insurance trends using social network such as Twitter

E-learning: Introduction to Data Visualization

40min – Beginner level

Abstract/Context

In this module we discuss about data visualization and why it becomes so crucial today to master this kind of technics. We highlight some common techniques of data visualization and give use cases for the insurance sector

Learning objectives

- Demonstrate why a relevant data visualization is important
- Present a set of data visualization technics
- Teach good and bad practices about data visualization
- Propose some insurance use cases

Agenda

1. Overview of data visualization
 - History of data visualization: an old subject
 - Reasons for Data visualization
 - Characteristics of accurate graphic representation : explore, explain a story, not distort and take into account the audience, keep it simple
 - Common issues
2. Data visualization families
 - Comparison:
 - Make the difference between values or observe similarities
 - Examples: bar chart, multiset bar chart, pyramid, radar,
 - Composition:
 - Highlight content of data (parts of data on its total)
 - Examples: pie chart, marimekko chart, treemap
 - Distribution:
 - Propose frequency and how data is spread out
 - Examples: box plot, density plot, histogram
 - Relationship:
 - Show connections or correlations between data
 - Examples: heatmap, network diagram, bubble chart,
3. Examples in Insurance
 - Dashboard for insurance reporting
 - Infographic studies
 - Other examples of data visualization in Insurance

E-learning: Emergence of new technologies

35min – Beginner level

Abstract/Context

In this module we present additional topics related to the data management in Insurance. In this module we mainly focus on three of them: Internet of things and its applications, chatbot technology and the interest for insurance distribution or management, and finally API systems.

Learning objectives

- Present new technologies related to data practices
- List pros and cons about these equipment
- Highlight how they can be useful for insurance sector

Agenda

1. IOT concept and application sectors
 - Definition of IoT and mapping of IoT initiatives
 - New image for insurance
 - New business model / pricing for insurer
 - Prevent/ manage risks in a better way
 - Risks and limits: data breach, ransomware, DDoS attack, etc.
2. Chatbot
 - What is a chatbot? Introduction with messaging app evolution, expert fields, definition and working process
 - Why a chatbot? Use of chatbot in industries, mimic human (first impression, evaluation)
 - How to create it?
 - Presentation of Facebook, Slack, API.ai systems.
 - Additional discussion around natural language processing
 - Some interesting applications in insurance
3. API development and functioning
 - Definition of an application - programming - interface (API)
 - How does it work?
 - Types of API : local, web, program
 - Aim of an API: example with and without API for weather application.
 - Interest for users, interest for developer and business.
 - Issues related to API: monetization, application shutdown, Security
 - Example of APIs (google map, google trad, Twitter)

E-learning: New waves in Insurance

35min – Beginner level

Abstract/Context

In this module we introduce the next trends about AI and new technologies within financial industries. We discuss more specifically on new actors and new technologies such as Blockchain and numerous benefits for insurance, with a specific use case about the smart contracts.

Learning objectives

- Have in mind next trends in insurance
- Understand market evolutions and new players
- Understand technological mechanism of Blockchain
- Demystify this technology with a simple and concrete business application

Agenda

1. Fintech, Insurtech and Startup AI environment
 - A rich ecosystem of numerous startups
 - A complex interaction with existing companies
 - Focus on Fintech: description, main clusters, examples of startups
 - Focus on Insurtech: description, main clusters, examples of startups
2. Blockchain technology and cryptocurrency
 - Definitions about Blockchain technology (main definitions, public vs private, rules)
 - How does it work? (process, type of distribution)
 - Relation with cryptocurrency (bitcoin, ethereum, ripple, etc.)
 - Pro/cons and misunderstanding
 - Blockchain potential applications
3. Application with smart contract
 - What is a smart contract?
 - Private space for customers
 - Sharing platform
 - Secure environment
 - What are the key benefits?
 - Single access point in a digital portal
 - Automatic control and automatic execution
 - Data quality and ownership
 - Costs
 - Application to the context of insurance underwriting

E-learning: Supervised machine learning part. 1 (advanced)

50min – Intermediate level

Abstract/Context

In this module we will focus on advanced machine learning techniques and how such models work practically. We will present a first set of supervised machine learning models including: decision trees, bagging and random forest models.

Learning objectives

- Have a global overview on the main machine learning techniques;
- Be able to compare methods between each other;
- Understand the key calculation steps behind models.

Agenda

1. Machine Learning context
 - 1.1 Introduction
 - 1.2 Artificial Intelligence origins and ML families
2. Error measure
 - 2.1 Introduction
 - 2.2 Classification problems: confusion matrix
 - 2.3 Classification problems: ROC curve
 - 2.4 Regression problems
 - 2.5 Overfitting
 - 2.6 Overfitting: training and test errors
 - 2.7 Cross-validation
3. Decision Tree
 - 3.2 Reminder
 - 3.3 Optimal split
 - 3.4 When to stop the splitting process - Notations
 - 3.5 Pruning
 - 3.6 Optimal tree
 - 3.7 Optimal tree
4. Bagging and random forest
 - 4.1 Introduction
 - 4.2 Bagging
 - 4.3 Random Forests
 - 4.4 Random Forests – Variable Importance
5. Conclusion

E-learning: Supervised machine learning part. 2 (advanced)

50min – Intermediate level

Abstract/Context

In this module we will focus on advanced machine learning techniques and how such models work practically. We will present a first set of supervised machine learning models including: Gradient boosting, Support vector machine, neural networks models.

Learning objectives

- Have a global overview on the main machine learning techniques;
- Be able to compare methods between each other;
- Understand the key calculation steps behind models.

Agenda

1. Introduction
2. Gradient boosted models
 - 2.1 Introduction
 - 2.2 Boosting trees
 - 2.3 Gradient descent
 - 2.4 Gradient boosting algorithm
 - 2.5 Features Importance
 - 2.6 Illustration
3. Neural networks
 - 3.1 Introduction
 - 3.2 Introduction to ANN
 - 3.3 The logistic regression
 - 3.4 Introduction to artificial neural networks
 - 3.5 Artificial neural networks : how it works
 - 3.6 Activation functions
 - 3.7 Backpropagation algorithm
 - 3.8 Illustration
4. Support vector machines
 - 4.1 Introduction
 - 4.2 Maximal margin classifier
 - 4.3 From maximal margin classifier to support vector machines
 - 4.4 From maximal margin classifier to support vector machines
 - 4.5 Illustration
5. Conclusion
6. References and further reading material

E-learning: Unsupervised machine learning (advanced)

50min – Intermediate level

Abstract/Context

In this module we will focus on advanced machine learning techniques and how such models work practically. We will present a first set of unsupervised machine learning models including: hierarchical ascending classification, k-means, dbscan.

Learning objectives

- Have a global overview on the main machine learning techniques;
- Be able to compare methods between each other;
- Understand the key calculation steps behind models.

Agenda

1. Introduction
 - 1.1 Introduction to unsupervised machine learning
 - 1.2 Families of machine learning models
2. Partitioning clustering
 - 2.1 Introduction
 - 2.2 Definition
 - 2.3 Distance measures
 - 2.4 Inertia
 - 2.5 K –Mean definition
 - 2.6 K –Mean algorithms
 - 2.7 K –Mean optimisation
 - 2.8 K –Mean sensitivity
3. Hierarchical clustering
 - 3.1 Introduction
 - 3.2 General concept
 - 3.3 Algorithm
 - 3.4: Example
4. Density-based clustering (DBSCAN)
 - 4.1 Introduction
 - 4.2 Presentation of DBSCAN
 - 4.3 Type of points
 - 4.4 Algorithm
 - 4.5 Sensitivity to parameters
5. Conclusion

E-learning: Advanced data visualization (advanced)

40min – Intermediate level

Abstract/Context

In this module we will study in details data visualization technics focusing on main process to define an accurate data visualization, but also on new technologies example and dashboarding good practices.

Learning objectives

- Be aware of the state of art techniques concerning data visualization
- Be able to create accurate data visualization according to the context of work
- Be aware of technologies and best practices about dashboarding

Agenda

1. Introduction
2. Process to set up a data visualization
 - 2.1 Introduction
 - 2.2 Prepare the data
 - 2.3 Questions to be asked and answered
 - 2.4 Data visualization improvement, example
3. How statistical approaches can improve data visualization
 - 3.1 Introduction
 - 3.2 Regression
 - 3.3 Correlation matrix
 - 3.4 Clustering
4. Business case, interpretation of telematics data
 - 4.1 Introduction
 - 4.2 Prepare the data
 - 4.3 Data visualization process
 - 4.4 Improvement of the model with statistical methods
5. Best practices
6. Conclusion
7. References and further reading material