Welcome to:

Machine Learning in Insurance Industry

Unit objectives

After completing this unit, you should be able to:

- Understand importance of machine learning in insurance
- Learn about the potential use cases of machine learning in insurance industry
- Understand the applications of machine learning in insurance climb analysis and pricing optimization
- Gain knowledge on applications of machine learning for personalized marketing in insurance industry
- Learn about machine learning for risk prediction in life insurance industry

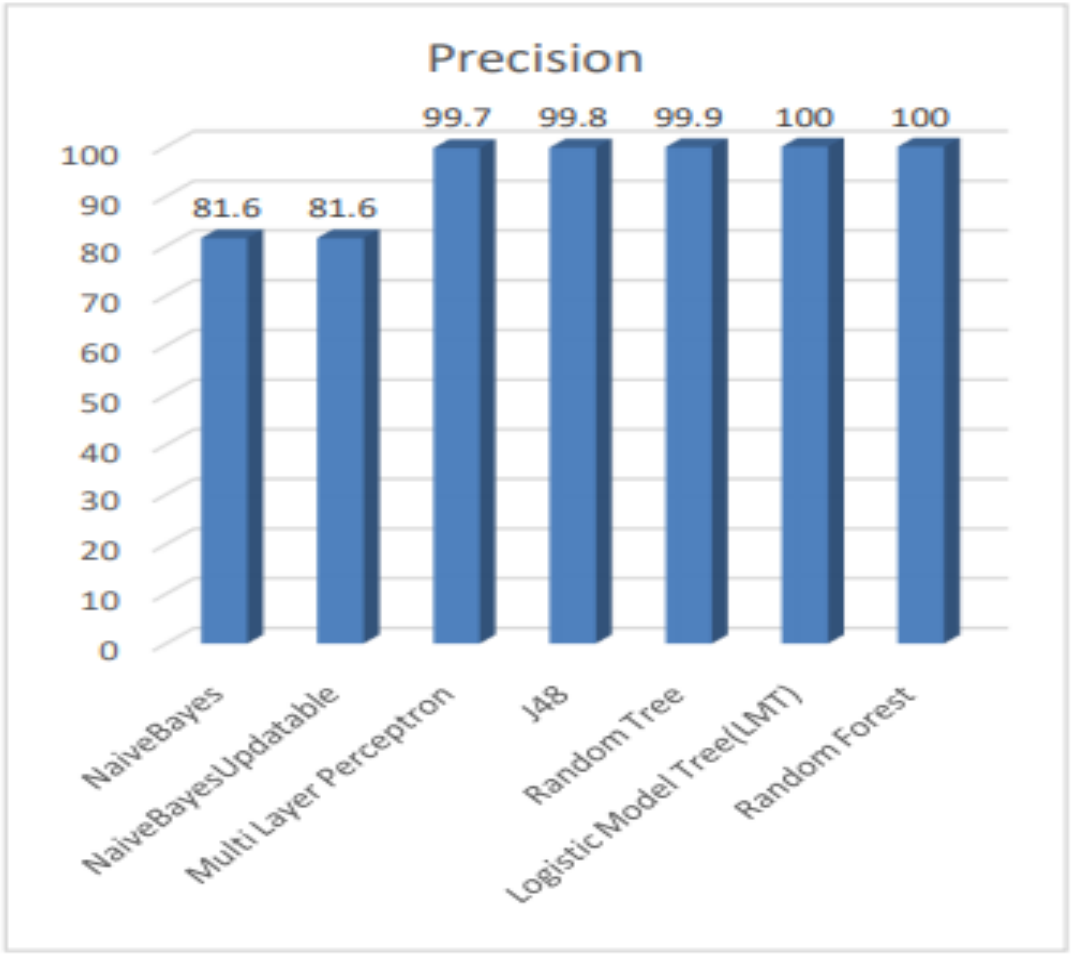
Importance of machine learning in insurance

- Importance of machine learning in insurance:
  - Revenue expansion.
  - Advisory excellence.
  - Improved operational efficiency.
  - Maximized customer experience.
  - Maximized customer experience.

Potential use cases of machine leaning in insurance industry

- Insurance Advice.
- Automation.
- Claim acceleration.
- Personalized marketing.
- Pricing sophistication.
- Risk identification and classification.
- Customer segmentation.
- Fraud prevention use case.

Case study on insurance climb analysis using machine learning algorithms



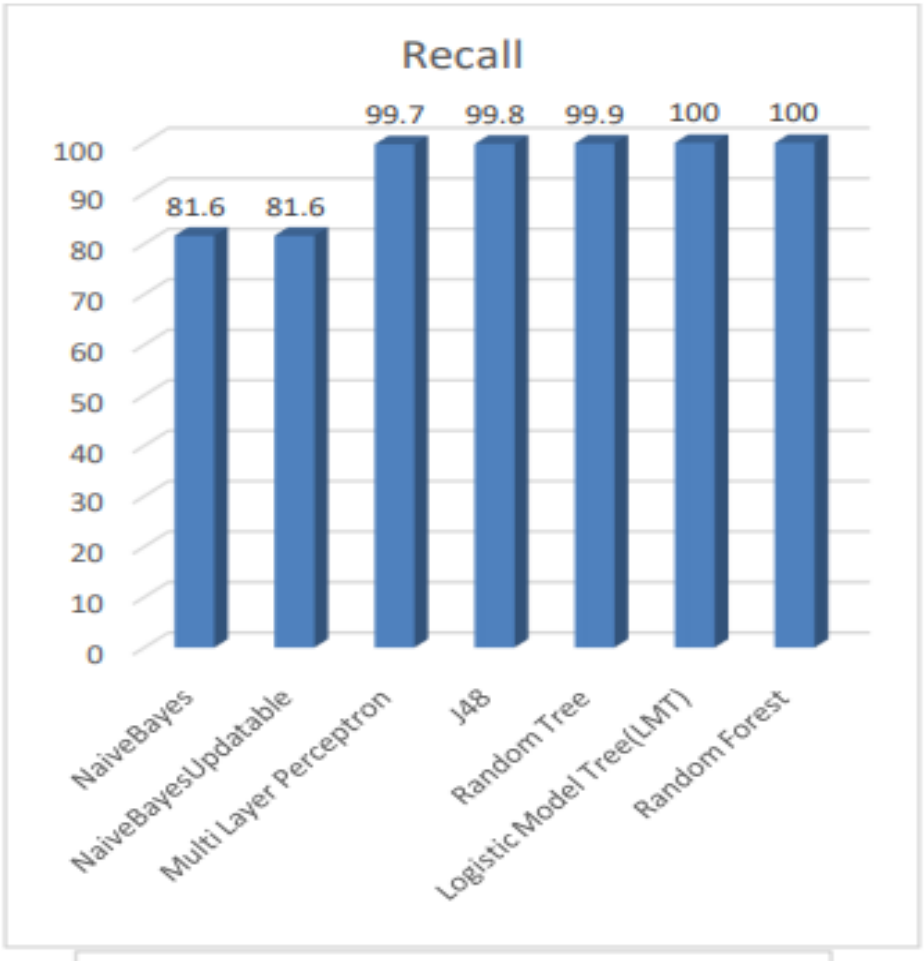


Figure: Insurance climb analysis using machine learning algorithms

Source: https://www.ijitee.org/wp-content/uploads/papers/v8i6s4/F11180486S419.pdf

Case study on using machine learning for insurance pricing optimization

- AXA, the large global insurance company, has used machine learning in a POC to optimize pricing by predicting “large-loss” traffic accidents with 78% accuracy.
- About 7-10 percent of AXA's clients each year trigger a car crash.
- Most of them are minor incidents costing hundreds or thousands of dollars in insurance premiums, but only 1% are named large-failure instances requiring over $10,000 in pay-outs.
- To maximize the price of its policies, it is critical for AXA appraisers to recognize which customers are at greater risk for such situations.

Personalized marketing in insurance industry

- How to create Impactful personalization:
  - Data: Identifying unmet needs.
  - Decisioning: Using customer behaviour to trigger effective campaigns.
  - Design: Finding the right mix to maximize effort.
  - Distribution: Test, pivot, invest.

Predictive model for insurance underwriting (1 of 3)

- The quality of this study determines the likelihood of failure or benefit for the organization adaptive analysis system is a computational approach that evaluates prior and current variables to include a rating that directs possible risks or prospects for the future.

Figure: Predictive model for insurance underwriting

Source: https://www.ijitee.org/wp-content/uploads/papers/v8i6s4/F11180486S419.pdf

Predictive model for insurance underwriting (2 of 3)

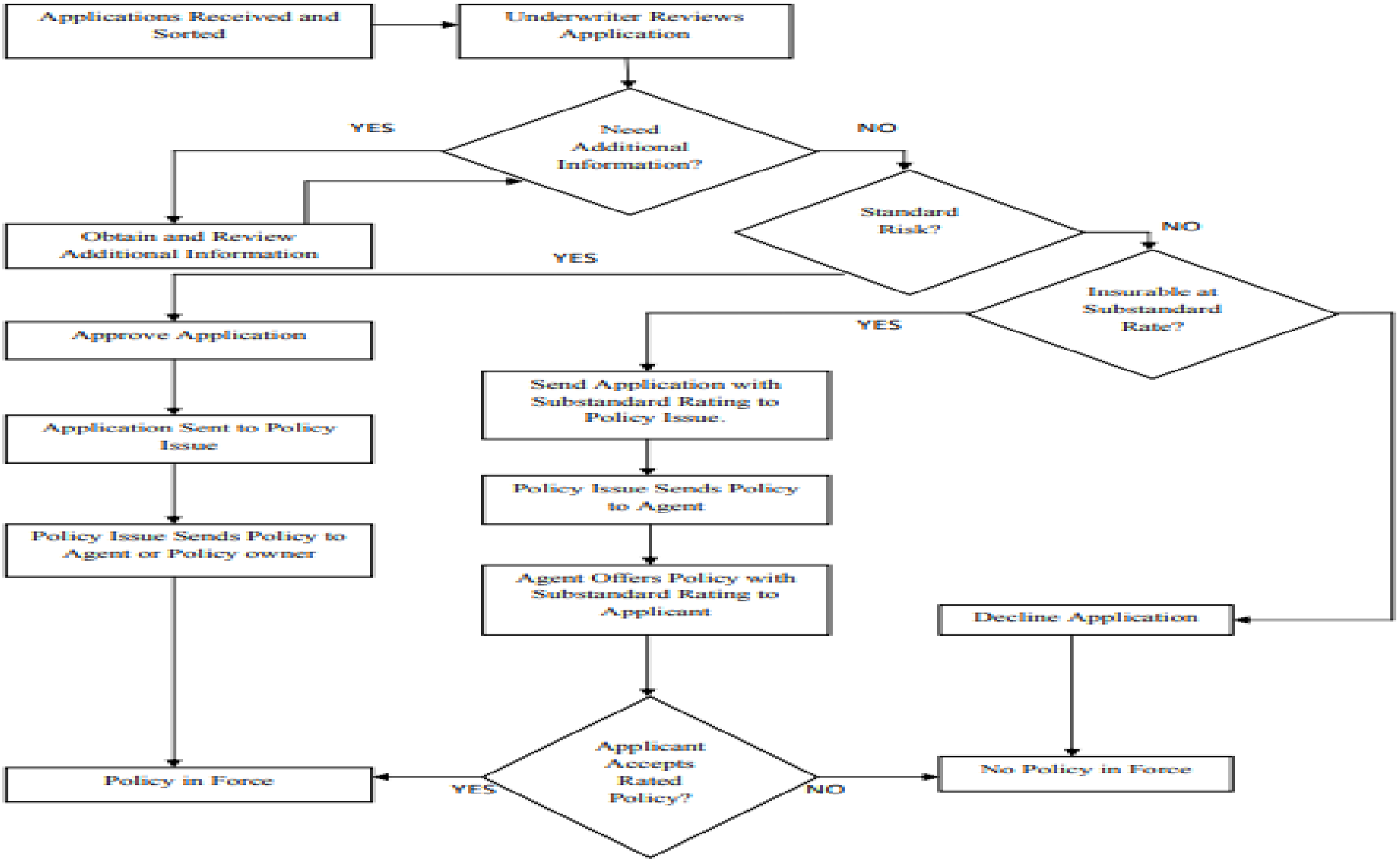


Figure: Underwriter Workflow

Source: https://www.ijitee.org/wp-content/uploads/papers/v8i6s4/F11180486S419.pdf

Predictive model for insurance underwriting (3 of 3)

- Anti-Selection.

- Anti-selection or unfavourable choice is the negative effect on insurers when choosing threats that are more likely to fail than the anticipated insurance rate.

- It can be quite costly for an insurance company which occurs when a gap in underwriting knowledge enables a greater threat category to purchase life or health insurance at the similar cost as a smaller threat category.

Case study: Risk prediction in life insurance industry (1 of 5)

- Life insurance agencies using supervised learning algorithms have gradually begun statistical analysis to enhance their company performance.

- Experts focused on information processing strategies to identify fraud between insurance companies, which due to large loss industries is a critical problem.

Case study: Risk prediction in life insurance industry (2 of 5)

- Table: Description of data set used.

Attributes	Type	Description
Product_Info_1-7	Categorical	7 normalized attributes concerning the product applied for
Ins_Age	Numeric	Normalized age of an applicant
Ht	Numeric	Normalized height of an applicant
Wt	Numeric	Normalized weight of an applicant
BMI	Numeric	Normalized Body Mass Index of an applicant
Employment_Info_1-6	Numeric	6 normalized attributes concerning employment history of an applicant
InsuredInfo_1-6	Numeric	6 normalized attributes offering information about an applicant
Insurance_History_1-9	Numeric	9 normalized attributes relating to the insurance history of an applicant
Family_Hist_1-5	Numeric	5 normalized attributes related to an applicant's family history
Medical_History_1-41	Numeric	41 normalized variables providing information on an applicant's medical history
Medical_Keyword_1-48	Numeric	48 dummy variables relating to the presence or absence of a medical keyword associated with the application
Response	Categorical	Target variable, which is an ordinal measure of risk level, having 8 levels

Case study: Risk prediction in life insurance industry (3 of 5)

- Comparison between correlation-based feature selection and principal components analysis feature extraction.
  - Supervised learning algorithms.
  - Multiple linear regression model.
  - Reptree algorithm.
  - Random tree.
  - Artificial neural network.

Case study: Risk prediction in life insurance industry (4 of 5)

- Experiments and results of case study.
- The attributes indicating more than 30% of the lost information would be removed from the study.
- Application information 1, application information 4, application information 6 and clinical record 1 are the only characteristics maintain for more study.
- The null explanation is that MCAR is the information that is lacking.

Case study: Risk prediction in life insurance industry (5 of 5)

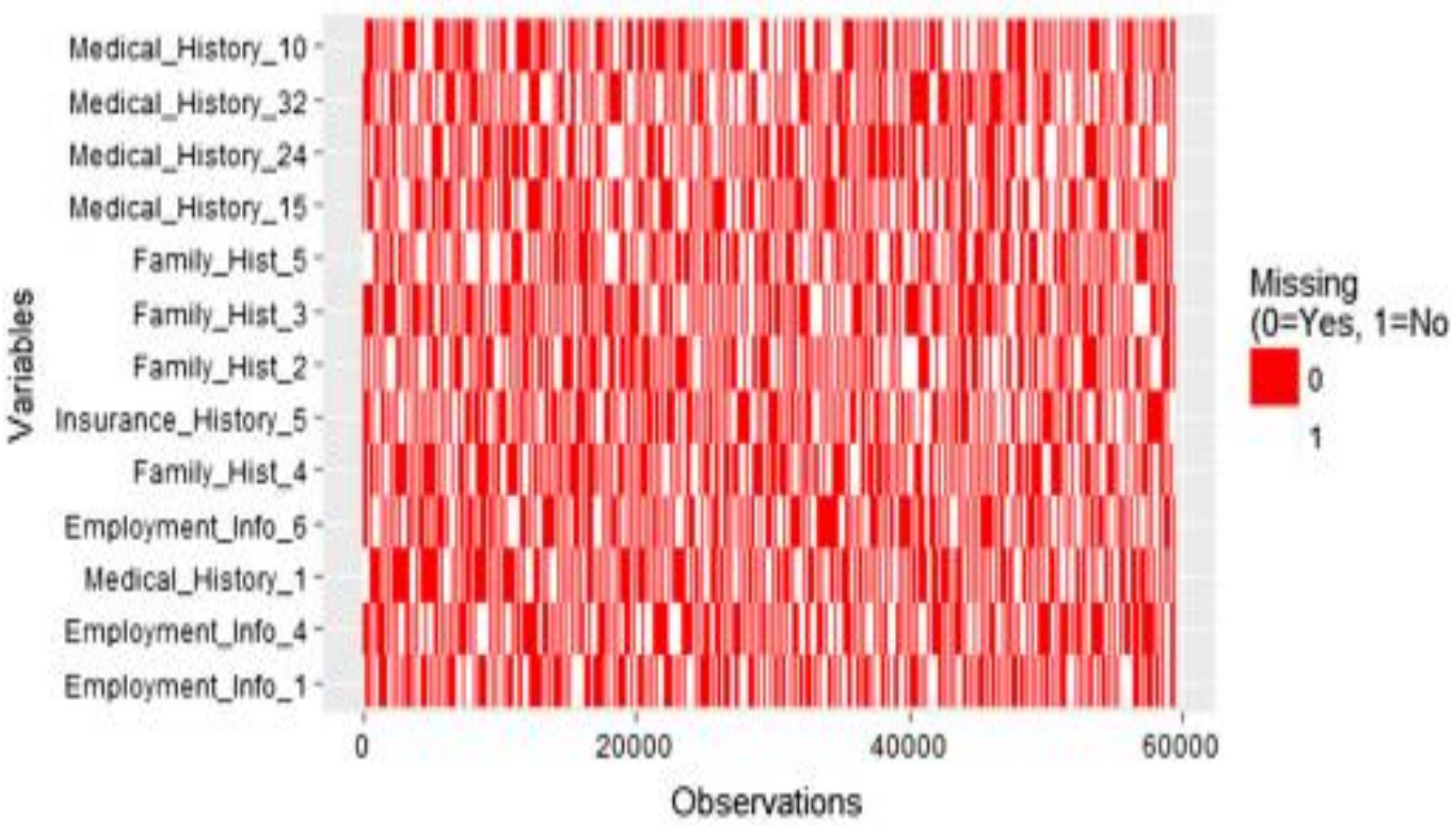


Figure: Missing value plot for train data

Source: https://www.ijitee.org/wp-content/uploads/papers/v8i6s4/F11180486S419.pdf

Checkpoint (1 of 2)

Multiple choice questions:

1. Which of the following is a good test dataset characteristic?
   a) Large enough to yield meaningful results
   b) Is a representative of the dataset as a whole
   c) Both A and B
   d) None of the above

2. Which of the following is a disadvantage of decision tree?
   a) Factor analysis
   b) Discussion tree are robust to outlier
   c) Decision tree are prone to be over fit
   d) None of the above

3. How do you handle missing or corrupted data in data.
   a) Drop missing rows or columns
   b) Replace missing values with mean/median/mode
   c) Assign a unique category to missing values
   d) All the above

Checkpoint solutions (1 of 2)

Multiple choice questions:

1. Which of the following is a good test dataset characteristic?

a) Large enough to yield meaningful results
b) Is a representative of the dataset as a whole
c) Both A and B
d) None of the above

2. Which of the following is a disadvantage of decision tree?

a) Factor analysis
b) Discussion tree are robust to outlier
c) Decision tree are prone to be over fit
d) None of the above

3. How do you handle missing or corrupted data in data.

a) Drop missing rows or columns
b) Replace missing values with mean/median/mode
c) Assign a unique category to missing values
d) All the above

Checkpoint (2 of 2)

Fill in the blanks:

1. PCA is an example for **----------** algorithm.
2. **----------** **a**lso known as one-to-one marketing or individual marketing, is a marketing strategy by which companies leverage data analysis and digital technology to deliver individualized messages and product offerings to current or prospective customers.
3. **----------** is a set of activities undertaken to prevent money or property from being obtained through false pretenses.
4. **----------** is the process of evaluating an insuranc**e** application that involves determining an applicant's risk by reviewing his/her medical information, financial information and lifestyle, and taking the applicant's age and gender into consideration.

Checkpoint solutions (2 of 2)

Fill in the blanks:

1. PCA is an example for **deterministic** algorithm.
2. **Personalized marketing,** also known as one-to-one marketing or individual marketing, is a marketing strategy by which companies leverage data analysis and digital technology to deliver individualized messages and product offerings to current or prospective customers.
3. **Fraud detection** is a set of activities undertaken to prevent money or property from being obtained through false pretenses.
4. **Underwriting** is the process of evaluating an **insurance** application that involves determining an applicant's risk by reviewing his/her medical information, financial information and lifestyle, and taking the applicant's age and gender into consideration.

Question bank

Two mark questions:

1. List the importance of Machine learning in Insurance Industry.
2. Why Machine learning is important in Insurance Industry?
3. What is personalized marketing ? How it can be used in Insurance industry?
4. What is insurance under writing?

Four mark questions:

1. List the importance of Machine learning in Insurance Industry
2. How Machine learning can be used in Insurance claim analysis?
3. How machine learning can be used for personalized marketing in insurance industry?
4. How machine learning can be used for insurance pricing optimization?

Eight mark questions:

1. Write a short notes on machine learning techniques used for insurance pricing optimization.
2. Explain how machine learning can be used for Risk prediction and classification in insurance industry.

Unit summary

Having completed this unit you should be able to:

- Understand importance of machine learning in insurance
- Learn about the potential use cases of machine learning in insurance industry
- Understand the applications of machine learning in insurance climb analysis and pricing optimization
- Gain knowledge on applications of machine learning for personalized marketing in insurance industry
- Learn about machine learning for risk prediction in life insurance industry