**Data Mining for Decision making**

**Assignment-1**

**Batch – 2022-24**

**MM.20 Date of Submission 25/12/2022**

**Q.1** Consider Geico, an auto insurance company. **Suppose Geico hypothetically plans to customize its auto insurance offerings and needs to understand what its customers view as important from their insurance provider.** Geico can ask its customers to rate how important the following two attributes are to them when considering the type of auto insurance, they would use:

* + savings on premium
  + existence of a neighborhood agent

The importance of the attributes is measured using a seven-point Likert-type scale, where a rating of one represents not important and seven represents very important.

Discuss the problem and suggest a data mining technique in detail for Geico for being competent in the market.

**Q.2** Data in spreadsheet is consist of youths who register for e-Kaushal platform. e-Kaushal assess these youths and recommend the suitable jobs. The objective is to suggest relevant training program to increase the employability.

* Observe the data and derive some key insights with explanation. (You may include graphs and charts)
* Can you derive some insights to increase employability if yes Explain.

1.

Here are some thoughts on how Geico can use the Likert-type scale to understand customer preferences for the two attributes it has identified:

1. Savings on premium: Geico can ask its customers to rate how important savings on premium is to them when considering the type of auto insurance they would use. The Likert-type scale can be used to ask a question such as "On a scale of 1 to 7, how important is it to you to save on your auto insurance premium?" Customers can select a rating that reflects their level of importance, with 1 indicating that savings on premium is not important to them and 7 indicating that it is very important.
2. Existence of a neighborhood agent: Geico can also ask its customers to rate how important it is to have a neighborhood agent when considering the type of auto insurance they would use. The Likert-type scale can be used to ask a question such as "On a scale of 1 to 7, how important is it to you to have a neighborhood agent for your auto insurance needs?" Customers can select a rating that reflects their level of importance, with 1 indicating that having a neighborhood agent is not important to them and 7 indicating that it is very important.

Once Geico has collected data on the importance ratings for these two attributes, it can analyze the results to understand how important each attribute is to its customers. For example, Geico can calculate the mean rating for each attribute and compare the two means to determine which attribute is more important to its customers. This information can be used to inform Geico's customization efforts and help it better meet the needs and preferences of its customers.

1.. The problem Geico is facing is to understand what attributes its customers consider important when choosing auto insurance. Specifically, Geico wants to determine the relative importance of two attributes: savings on premium and the existence of a neighborhood agent. To do this, Geico can use a data mining technique called conjoint analysis.

Conjoint analysis is a statistical technique used to determine how customers value different attributes of a product or service. It involves presenting customers with a set of hypothetical products or services that vary in terms of their attributes (e.g., price, features, etc.) and asking them to choose which product or service they would prefer. By analyzing the choices made by customers, conjoint analysis can estimate the relative importance of each attribute.

To apply conjoint analysis to Geico's problem, Geico can create a set of hypothetical auto insurance policies that vary in terms of savings on premium and the existence of a neighborhood agent. For example, Geico could create four hypothetical policies:

1. High savings, no neighborhood agent
2. High savings, with neighborhood agent
3. Low savings, no neighborhood agent
4. Low savings, with neighborhood agent

Geico can then ask its customers to rate how likely they would be to choose each policy on a seven-point Likert-type scale. By analyzing the ratings, conjoint analysis can estimate the relative importance of savings on premium and the existence of a neighborhood agent.

In addition to estimating the relative importance of each attribute, conjoint analysis can also be used to determine the optimal level of each attribute. For example, Geico could use conjoint analysis to determine the optimal level of savings on premium that maximizes customer satisfaction while still maintaining profitability.

Overall, conjoint analysis can provide valuable insights into what attributes are important to Geico's customers and how Geico can tailor its auto insurance offerings to meet their needs.

2.1

* 2. Suggestions to Increase Employability:
* Based on the insights gained from analyzing the data, e-Kaushal could develop targeted training programs that address the most pressing skills gaps among the youth population. For example, if the data shows that many youths lack basic computer skills, e-Kaushal could develop a training program that focuses on teaching these skills.
* e-Kaushal could also partner with employers to develop training programs that are tailored to the specific needs of their industry. For example, if the data shows that there is a high demand for skilled workers in the healthcare industry, e-Kaushal could partner with healthcare employers to develop a training program that prepares youth for these types of jobs.
* Additionally, e-Kaushal could provide incentives for youths to complete their training programs, such as job placement services or financial rewards. This could help to motivate youths to take their training seriously and increase their chances of finding employment after completing the program.

This dataset contains information about candidates who have applied for vocational training in different sectors. Each row represents a unique candidate and includes information such as their age, gender, education, technical education, and assessment results. The dataset is structured in 19 columns with 23 rows.

Here is a brief explanation of each column:

* C.candidateId: Unique identifier for each candidate
* Der.Age: Age of the candidate
* C.personalDetails.gender: Gender of the candidate
* C.personalDetails.differentlyAbled: Indicates whether the candidate is differently-abled or not
* C.personalDetails.maritialStatus: Marital status of the candidate
* C.contactDetails.district: District of the candidate
* C.PreTrainingStatus: Indicates whether the candidate had any prior training or not
* Der.Experience: Indicates the experience level of the candidate
* EduD.education: Indicates the highest level of education attained by the candidate
* C.TechnicalEducation: Indicates whether the candidate has any technical education or not
* B.sector.name: Indicates the sector for which the candidate is applying
* B.batchfees: Indicates the fees for the training batch
* AJD.jobName: Indicates the job name
* QPS.qpParamOne.paramDesc: Indicates the job description
* edu.Per: Indicates the percentage of marks scored by the candidate in their highest level of education
* assmtforjob.result: Indicates the assessment result of the candidate (PASS/FAIL).

Key Insigths

1. Demographic Analysis: You could analyze the demographics of the youths registered on the e-Kaushal platform. This could include their age, gender, location, and education level. This information can help identify which groups are most interested in the platform and may also suggest where there are gaps in the market.
2. Skill Analysis: You could analyze the skills and interests of the youths registered on the e-Kaushal platform. This could help identify which industries and sectors are most in demand and where there are opportunities for growth.
3. Training Program Analysis: You could analyze the training programs offered on the e-Kaushal platform. This could include the duration of the program, the type of program, and the success rate of the program in terms of job placement. This information can help identify which programs are most effective and where there are opportunities for improvement.
4. Job Placement Analysis: You could analyze the job placements of the youths who have completed the training programs on the e-Kaushal platform. This could include the type of job, the location of the job, and the starting salary. This information can help identify which industries and sectors are most in demand and which skills are most valuable in the job market.
5. Trend Analysis: You could analyze the trend of the data over time to identify any patterns or changes in the market. This could include changes in demographics, changes in the demand for skills, or changes in the job market. This information can help identify future opportunities and challenges.

Graphs and charts can be used to represent the data in a more visual way, making it easier to analyze and draw insights. For example, a bar chart could be used to represent the demographics of the youths registered on the e-Kaushal platform, while a line chart could be used to show the trend of the data over time.