RESEARCH METHODOLOGY

OUR TEAM

SHARMEEN SHAIKH
86092300016

REESE PEREIRA
86092300007

PURVA BURUNDKAR
86092300018

BREEMA ALIAS
86092300010

PROBLEM STATEMENT

HEALTH INSURANCE PREDICTION AND ANALYSIS BY ARTIFICIAL NEURAL NETWORKS

As uninsured persons have faced huge medical expenses as a result of Covid, there was a significant demand for health insurance policies, raising awareness and the necessity for coverage. The research hopes to provide insights for healthcare stakeholders such as insurance companies, healthcare providers, and patients.

Problem

UNPREDICTABLE LIFE

As a result, in todays uncertain and stressful life, it is better to be insured ourselves and family.

SUFFICIENT COVERAGE

Coverage of medical and health policies should be sufficient enough covering all major disease.

CASHLESS TREATMENT

Choosing such insurance carriers, where cashless treatments are provided in all major hospitals, is advantageous.

TYPES OF RESEARCH

01

QUALITATIVE

focuses on comprehending the context, viewpoints, and subjective elements of using neural networks to predict health insurance.

02

DESCRIPTIVE

aims to provide a comprehensive overview of the current state, characteristics, and trends related to the application of neural networks in predicting health insurance outcomes.

03

APPLIED

delves into the challenges faced by individuals, organizations, and governments due to fraudulent claims, privacy breaches, and security vulnerabilities.

TYPES OF RESEARCH

04

QUANTITATIVE

involves using numerical data and Machine Learning Model to analyse study the effectiveness, accuracy, and performance of neural network models in predicting health insurance outcomes.

05

Analytical

involves scrutinizing data and theories to understand issues, offering evidence-based insights for better decision-making, optimizing health insurance claims, and improving the industry.

LITERATURE REVIEW

CITATION-1 (SHARMEEN SHAIKH-86092300016)

“Kapadiya, K., Patel, U., Gupta, R., Alshehri, M. D., Tanwar, S., Sharma, G., & Bokoro, P. N. (2022). Blockchain and AI-Empowered Healthcare Insurance Fraud Detection: An Analysis, Architecture, and Future Prospects. *IEEE Access*, *10*, 79606-79627. https://ieeexplore.ieee.org/abstract/document/9843995”

DISCUSSIONS

In view of the growing number of health problems, the author discusses the importance of health insurance while noting fraud, security, and other privacy issues.

FOCUS

focuses on the negative effects of health insurance fraud and highlights the demand for fraud detection technologies. provides a thorough analysis of the use of AI and blockchain for safe health insurance fraud detection while at the same time presenting a taxonomy for categorizing security issues.

RESEARCH GAP

there may be a gap in research regarding the practical challenges and real-world implementation of such systems.

RESULTS

paper's contributions are emphasized in the conclusion, as are the difficulties in implementing the suggested method in practical settings.

CITATION-2 (SHARMEEN SHAIKH-86092300016)

“Ho, C. W. L., Ali, J., & Caals, K. (2020). Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance. *Bulletin of the World Health Organization*, *98*(4), 263–269. https://doi.org/10.2471/BLT.19.234732”

HIGHLIGHTS

warns against improper use and calls for collaborative relationships between insurers, regulators, and insureds to maintain trust and sustainability.

FOCUS

talks about ethical and regulatory environment for data analytics in health insurance, while at the same time focusing on major aspects for it like effective data governance framework, a clear and accountable process

RESEARCH GAP

potential research gap could be to investigate in more detail the privacy and ethical concerns that arise when implementing AI and big data analytics in the health insurance industry.

RESULTS

conclusion highlights the potential of big data technologies for insurers, improving sustainability reporting, loss mitigation, and health insurance design.

BREEMA ALIAS SAP ID: 86092300010

Bhardwaj, N., & Anand, R. (2020). Health insurance amount prediction. *Int. J. Eng. Res*, *9*, 1008-1011

ABSTRACT

.
The research aimed to help individuals determine their health-specific funding needs. It focused on comparing three regression models: Multiple Linear Regression, Decision Tree Regression, and Gradient Boosting Decision Tree Regression for algorithm performance analysis.

CONCLUSION

Age and smoking status significantly influence predictions across all algorithms. The Gradient Boosting Regression model, based on decision trees, emerged as the top-performing model.

LIMITATIONS

- This research focuses on people to get an idea about the necessary amount required according to their health status rather than other company’s insurance terms and conditions.

- Advanced machine learning techniques are not used for the prediction.

BREEMA ALIAS SAP ID: 86092300010

Goundar, S., Prakash, S., Sadal, P., & Bhardwaj, A. (2020). Health Insurance Claim Prediction Using Artificial Neural Networks. *International Journal of System Dynamics Applications (IJSDA)*, *9*(3), 40-57.

ABSTRACT

The author introduces an artificial neural network model for predicting annual medical claims. After implementing the model, the author aimed to reduce the mean absolute percentage error by fine-tuning parameters like epochs, learning rate, and neuron count in various layers.

CONCLUSION

He concludes that the ANN model outperformed the human prediction that is used now. The ANN model reduced the error rate by about 11.5%.

LIMITATIONS

- Research was based on a particular company and author doesn't mentioned about any specific parameters that affected health insurance claim prediction.

- Also the author mentioned that the Training dataset was large which leads to overtraining which probably affected test data set.

PURVA BURUNDKAR SAP ID:86092300018

Christina X Ji, A. M. (2023, June). *Large-Scale Study of Temporal Shift in Health Insurance Claims*. Retrieved from Proceedings of Machine Learning Research: https://proceedings.mlr.press/v209/ji23a

MAIN IDEA

Building an algorithm to test for temporal shift in health insurance claim either at the population level or within a discovered sub-population

METHODS

This algorithm enables author to perform the first comprehensive evaluation of temporal shift in healthcare to our knowledge. The algorithms are,

- Hypothesis Testing
- Discovering Sub-populations with Shifts.
- Algorithm to Test for Temporal Shift.
- Meta-Algorithm to Scan for Temporal Shift

CONCLUSION

The study emphasizes the prevalence of temporal shifts in healthcare settings and suggests future work involving subgroup discovery methods and model updates to mitigate their impact on patient care.

CITATION-1 (REESE PEREIRA-8609230007)

"Kaushik, K., Bhardwaj, A., Dwivedi, A. D., & Singh, R. *Machine Learning-Based Regression Framework to Predict Health Insurance Premiums*. Retrieved from MDPI: https://www.mdpi.com/1660-4601/19/13/7898"

MAIN IDEA

- The authors predicted the health insurance cost incurred by individuals on the basis of their features.
- On the basis of various parameters, such as age, gender, BMI, number of children, smoking habits, and geolocation, an artificial neural network model was trained and evaluated.
- The experimental results displayed an accuracy of 92.72%, and the authors analysed the model's performance using key performance metrics.
- This research resulted that the impact of machine learning on health insurance will save time and money for both policyholders and insurers.

RESEARCH GAP

Exploring methods to ensure that sensitive health-related information is used ethically and securely in the predictive modeling process, considering maintaining privacy.

CITATION-2(REESE PEREIRA-8609230007)

“Matloob, I. K. Khan, S. A., Hussain, F., Butt, W. H., Rukaiya, R., & Khalique, F. (2021, September 13). *Open AccessArticle*. Retrieved from MDPI: https://www.mdpi.com/2076-3417/11/18/8478”

MAIN IDEA

- This paper presents a novel methodology based on machine learning to optimize medicalbenefits in healthcare sector.

- Our proposed methodology generated need-based packages using a machine learning model based on K means clustering. With the help of this model, we have computed the optimum premium amount.

- The results indicate that the medical premium amount is optimized by 25% of the current benefit amounts. Therefore, if adopted, it will not only allow employers and insurance companies to design suitable insurance schemes for the provision of healthcare benefits but will also prevent financial losses in the long run.

RESEARCH GAP

Designing strategies to incorporate user feedback and preferences into the clustering process, allowing individuals to have an active role in shaping their insurance packages.

HYPOTHESIS TESTING

1. Role of AI In Analysis of Insurance Claims Based on Health

Null Hypothesis (*H*0):

AI does not significantly improve the accuracy of health insurance prediction compared to traditional method

Alternative Hypothesis (*H*a):

AI significantly improves the accuracy of health insurance prediction compared to traditional methods.

2. Enhancing Disease Risk Prediction with ANNs

Null Hypothesis (*H*0):

ANNs will underperform traditional risk assessment models in terms of sensitivity and specificity.

Alternative Hypothesis (*H*a):

ANNs can outperform traditional risk assessment models in terms of sensitivity and specificity.

3. Impact of Lifestyle Factors In Claim Prediction

Null Hypothesis (*H*0):

Lifestyle factors (e.g., smoking, exercise, diet) has no effect on the types and costs of health insurance claims.

Alternative Hypothesis (*H*a):

Lifestyle factors (e.g., smoking, exercise, diet) has an impact on the types and costs of health insurance claims.

4. Geographic Variations in Healthcare Costs

Null Hypothesis (*H*0):

There are no significant geographic variations in healthcare costs.

Alternative Hypothesis (*H*a):

There are significant geographic variations in healthcare costs.

5. Demographics and Health Outcomes

Null Hypothesis (*H*0):

There are no differences in health outcomes among various demographic groups (e.g. age, gender, ethnicity) based on health insurance claims.

Alternative Hypothesis (*H*a):

There are differences in health outcomes among various demographic groups (e.g. age, gender, ethnicity) based on health insurance claims.

SAMPLE DATA FOR TESTING

- The data for this particular research based on health insurance will be provided by an insurance company(Primary Data would be provided).
- Since research involves confidential data it isn't feasible to collect real-world data, so we can utilize pre-existing data based on known distributions or patterns. It can be taken from websites like (IRDAI, Kaggle).
- Existing academic studies related to health insurance can provide valuable insights and data that we can reference or build upon in our research.

RESEARCH DESIGN

WHAT IS IT?

- A blueprint for a research study.
- Detailed plan that outlines what you want to study, how you'll collect data, and how you'll analyze it.
- Also be thought of as a roadmap that helps you find answers to your research questions by guiding every step of your investigation, from the beginning to the end.
- Crucial for ensuring your research is organized, valid, and reliable.

SAMPLING DESIGN

PURPOSE:

is to obtain a representative subset of data from a larger population of health insurance claims

SAMPLING SIZE:

Collected the data from 40 individuals with different health insurance claims

SAMPLING FRAME:

Questionnaires were circulated to collect data from different group of people

SAMPLING TECHNIQUE:

Stratified sampling involves dividing the population into subgroups or strata based on specific characteristics (e.g., age, gender, geographic location)..

OBSERVATIONAL DESIGN

PURPOSE:

- Observational research provides insights into real-world behaviors and phenomena.

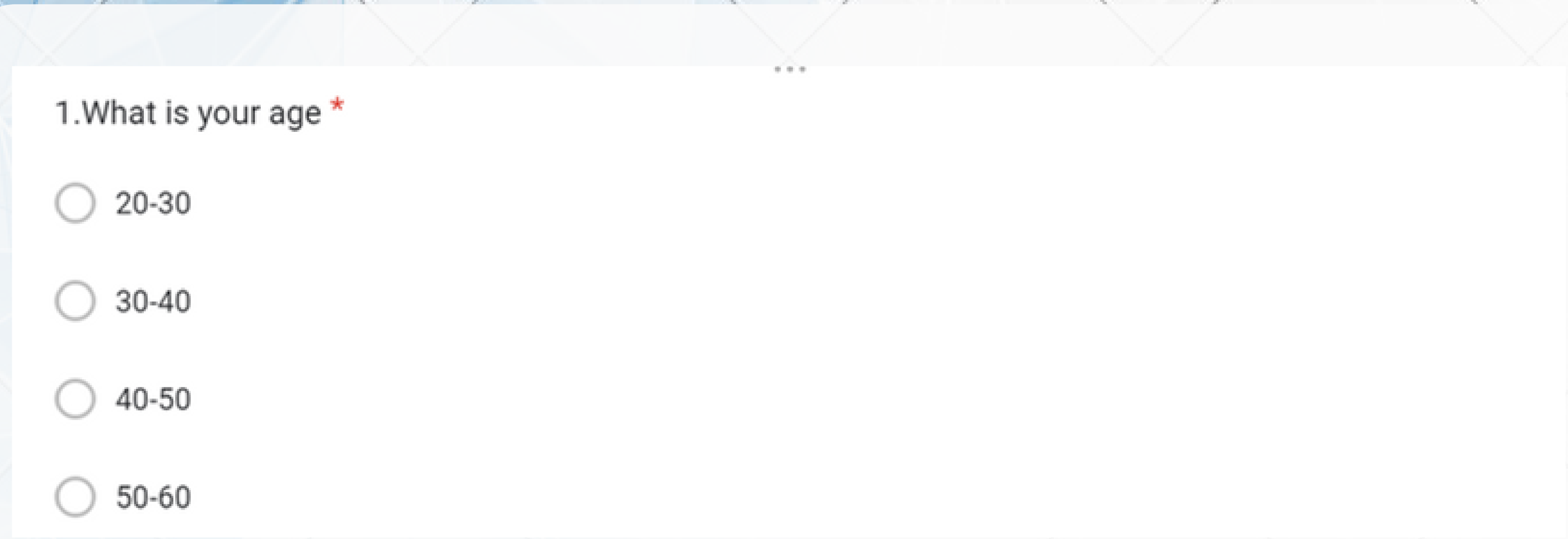
- This method aids in identifying patterns and trends.

ANALYSIS

- ANN is applied to the dataset to analyze and predict claims

- Thematic Content Analysis is a valuable tool for uncovering the hidden patterns and themes that influence health insurance claims, ultimately enhancing the accuracy of predictive models.

PILOT SURVEY



In order to know the variance in ages of different individuals we collected information about age groups.

2.What is your gender? *

- Male
- Female
- Others

To know the impact of gender in receiving claims.

3.Do you have diabetes? *

- Yes
- No

As having diabetes is proven to be the most chronic condition in impacting a persons health.

4.How many children you have? *

- 0
- 1
- 2
- 3
- 3+

Greater number of dependents increases chances of policy holder issuing claims to an individual.

5.What type of health insurance coverage do you have? *

- Medical insurance
- Dental insurance
- Vision insurance
- Life insurance
- Other

To get insights about different kinds of health insurance that people apply for

6.Have you ever received medical advise/treatment related to smoking or alcohol consumption?

- ○ No
- ○ Yes

In order to get details about smoking/alcohol consumption character of the policyholder

7. What is your current place of residence? *

- North
- South
- East
- West
- Central

To analyzing claims based on geographical regions.

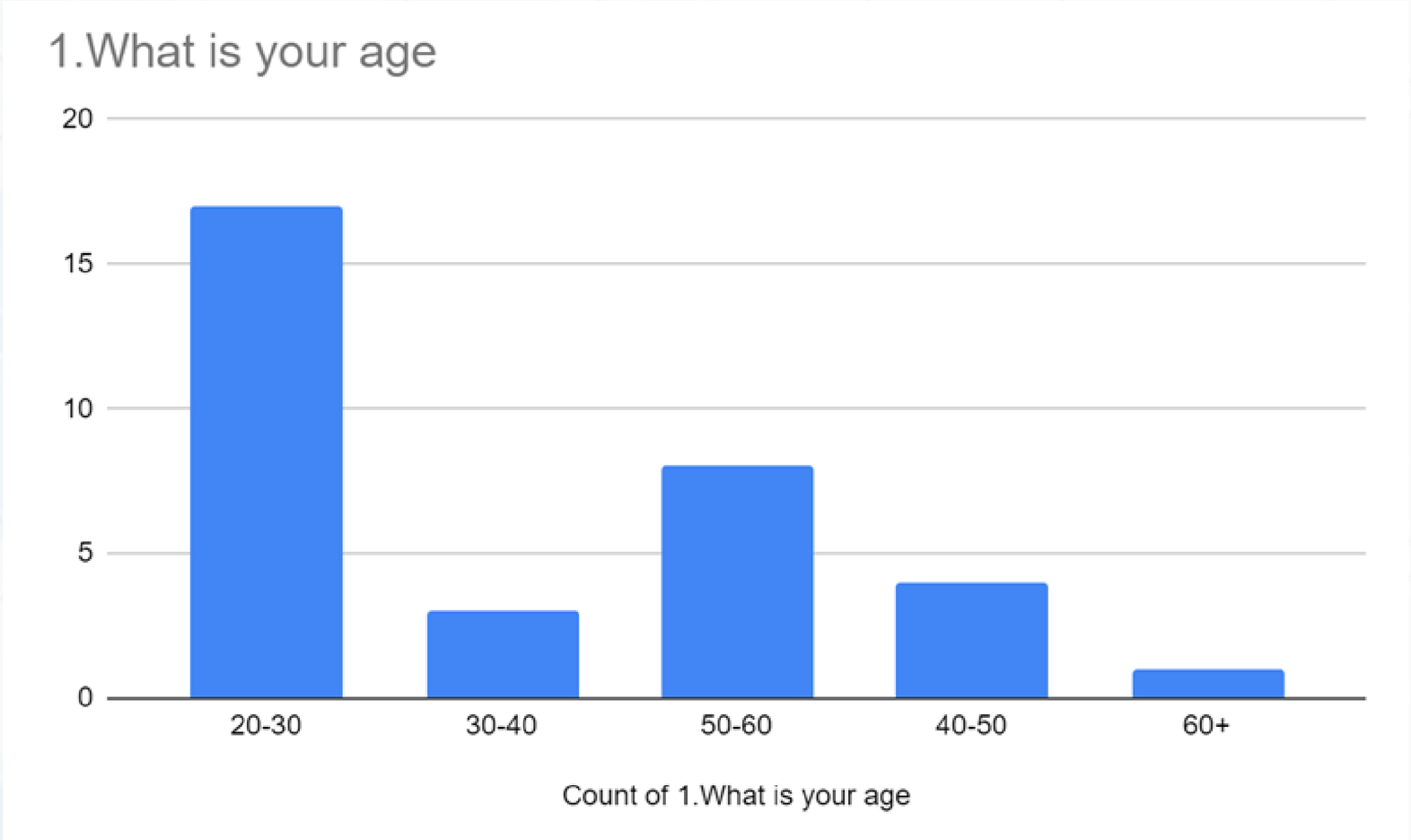
8. In your opinion, what is the most significant benefit of incorporating AI into health insurance claims processing? *

- Faster claims processing and approval
- Improved accuracy in claims assessment
- Enhanced fraud detection and prevention
- Better customer support and communication

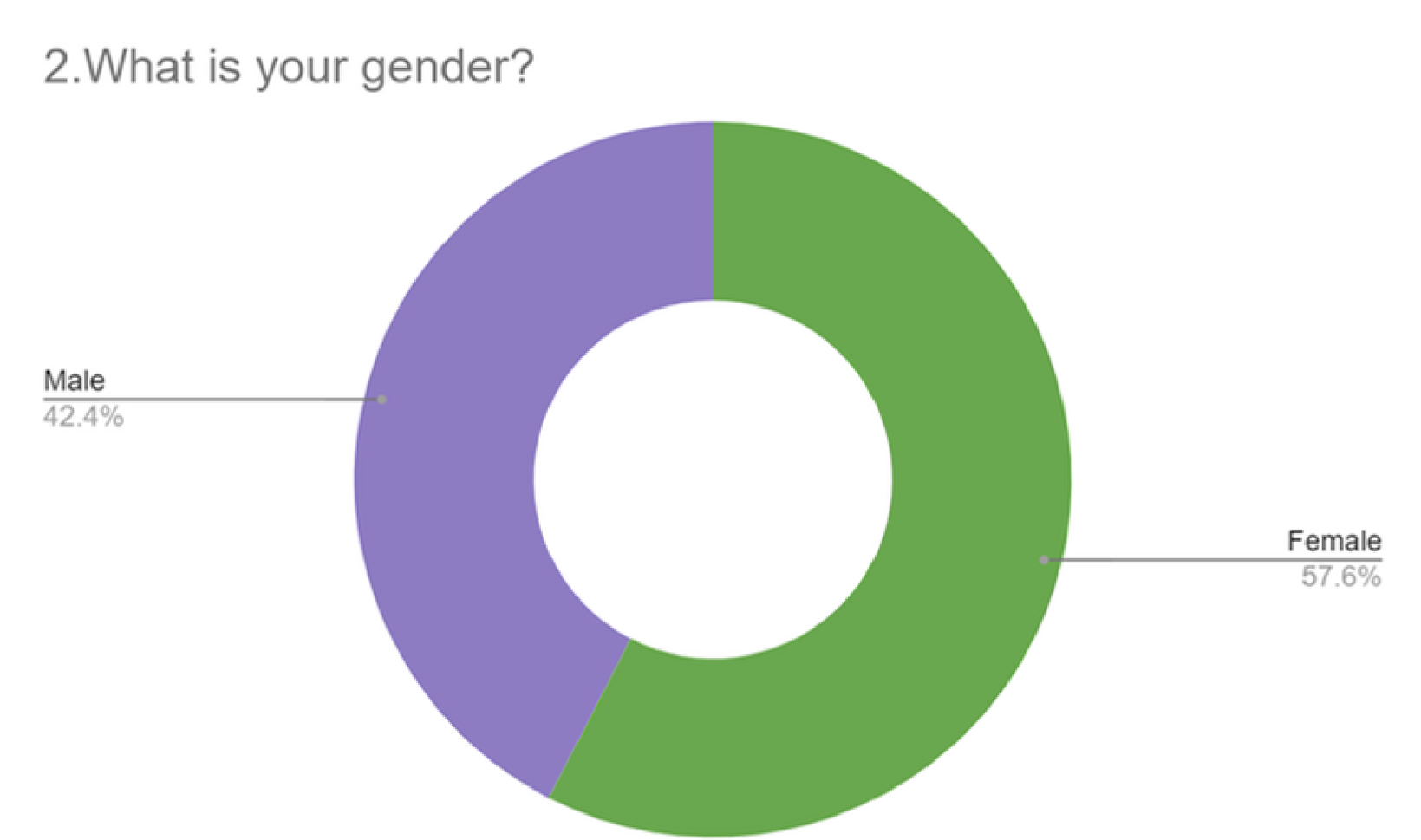
To analyse opinions of people about implimenting AI in health insuarance claim prediction.

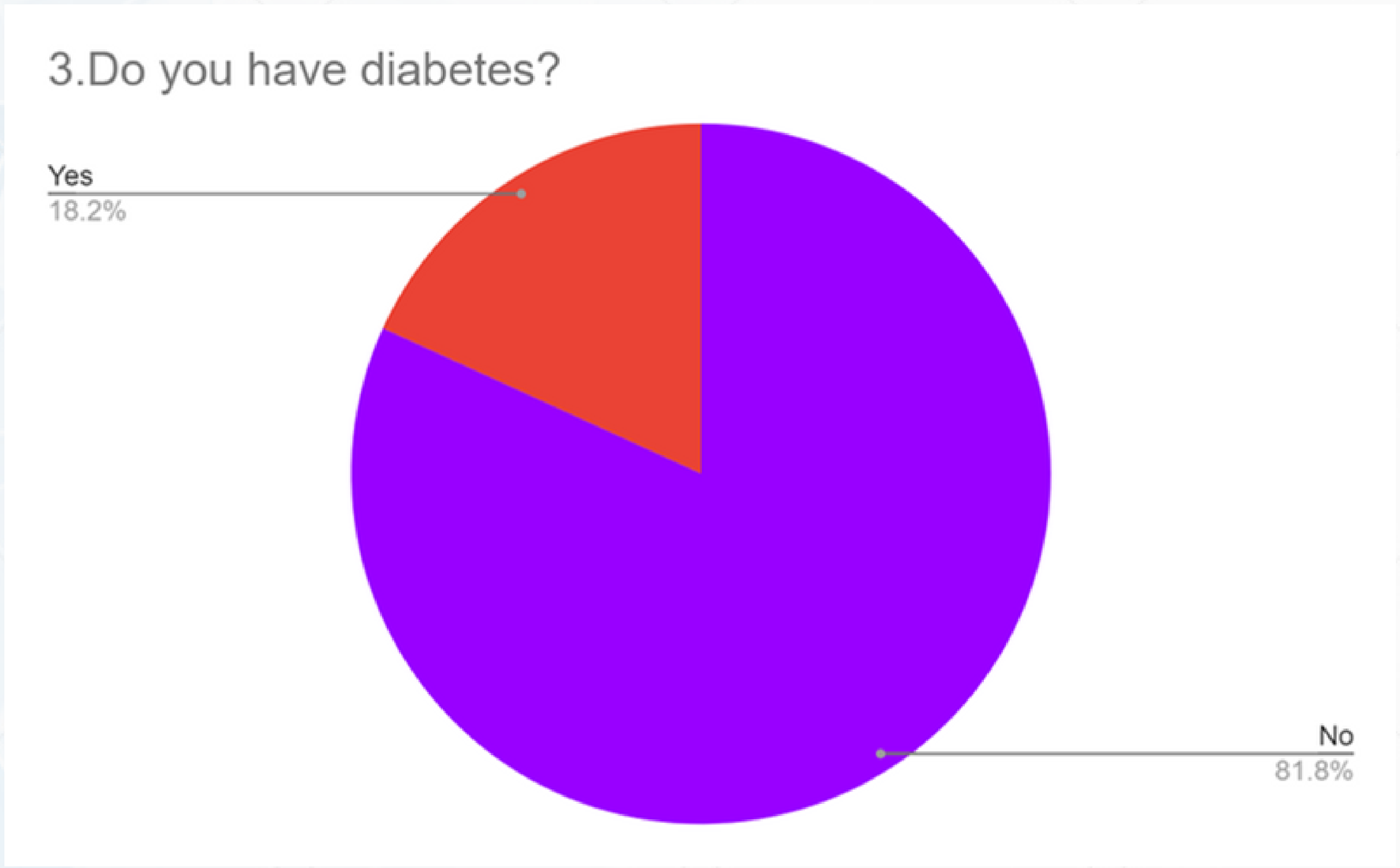
PILOT DATA SUMMARY

Age group

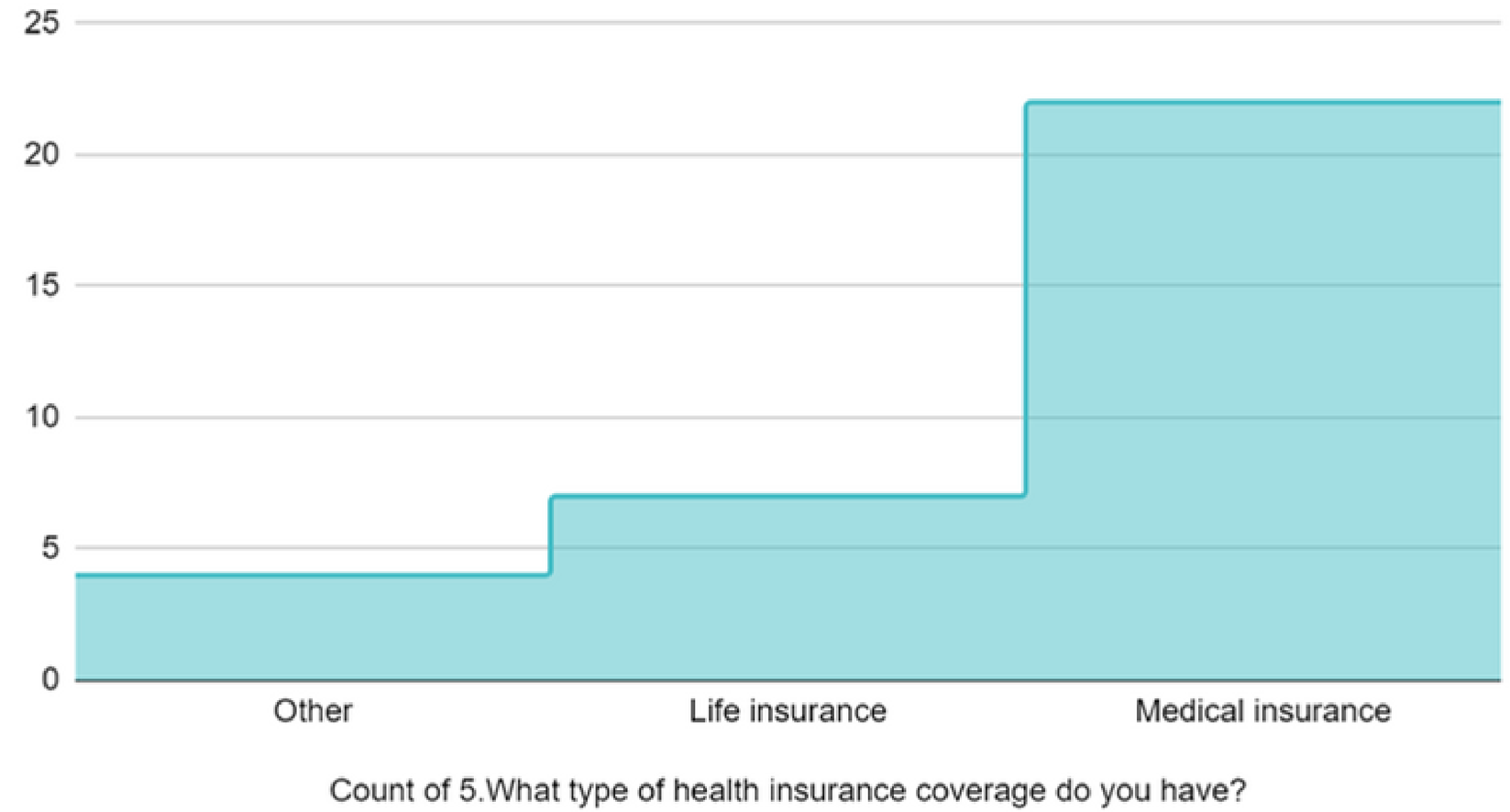

1.What is your age
20
15
10
5
0
20-30
30-40
50-60
40-50
60+
Count of 1.What is your age

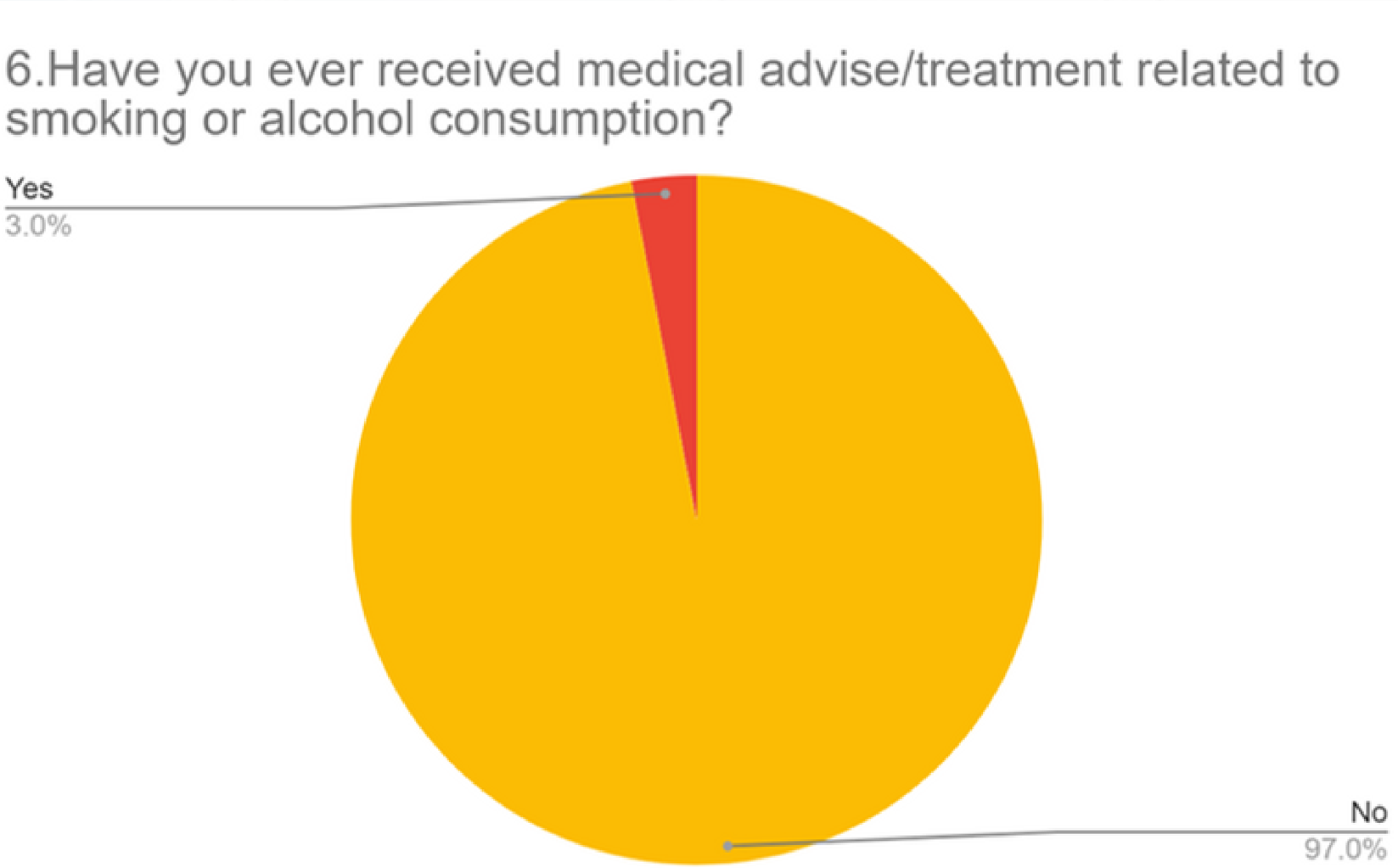
Gender

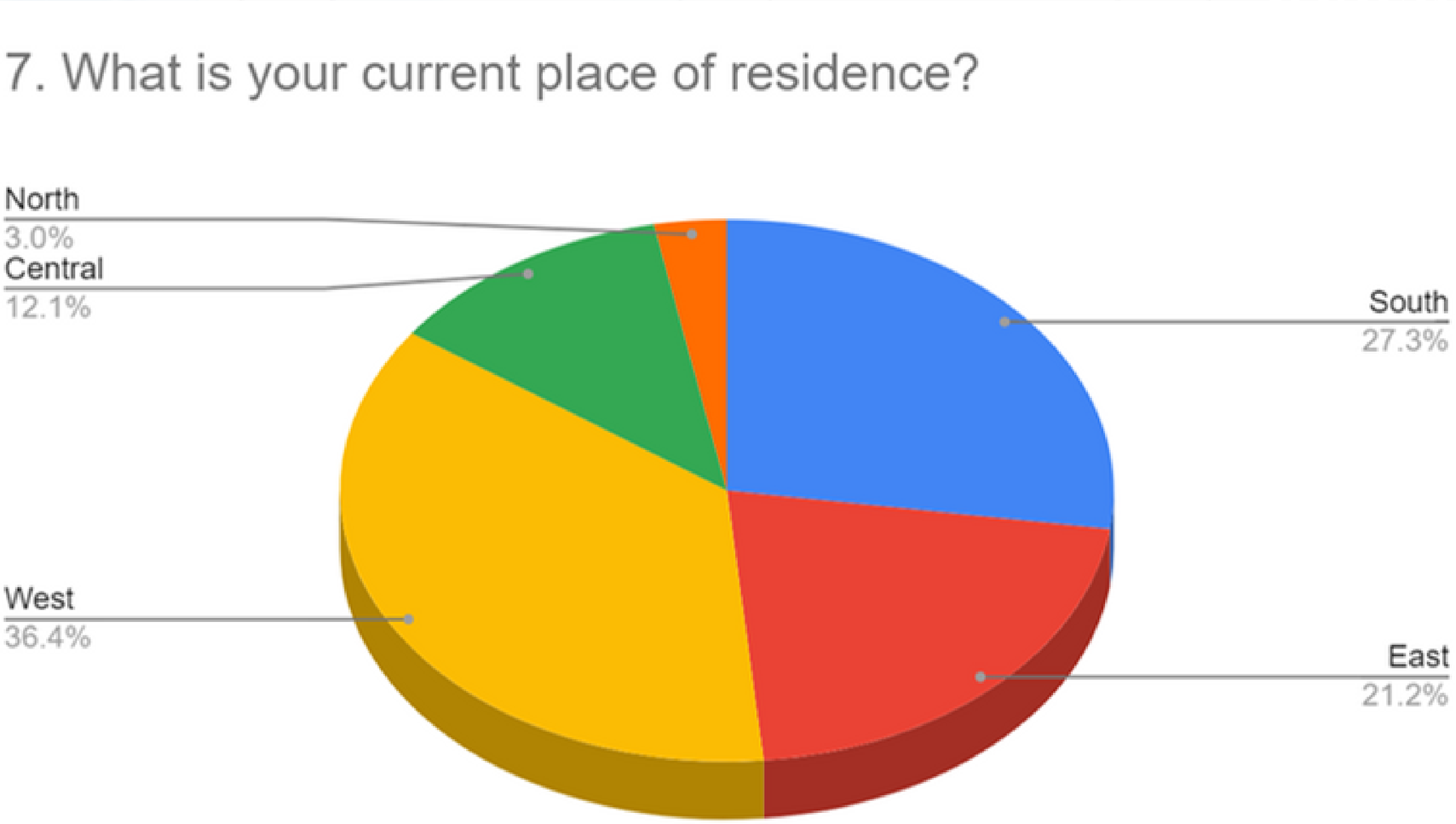



3.Do you have diabetes?
Yes
18.2%
No
81.8%

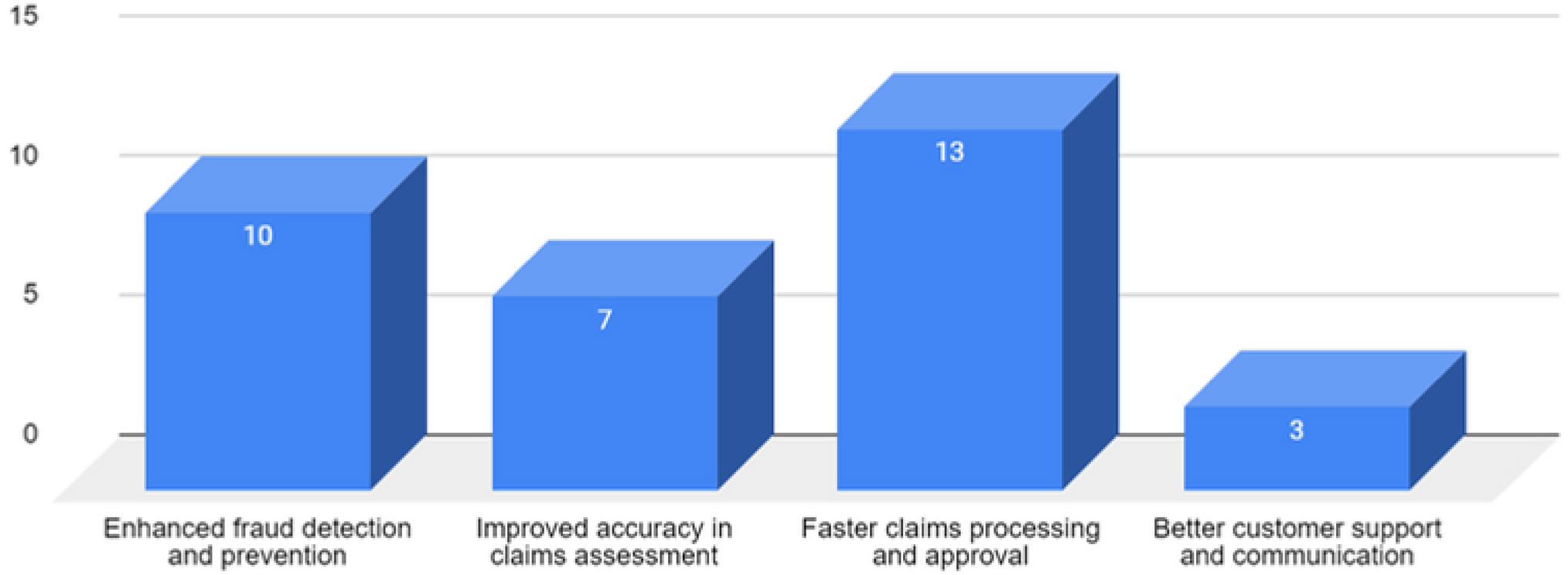
No of dependents


5.What type of health insurance coverage do you have?
25
20
15
10
5
0
Other
Life insurance
Medical insurance
Count of 5.What type of health insurance coverage do you have?


6.Have you ever received medical advise/treatment related to smoking or alcohol consumption?
Yes
3.0%
No
97.0%


7. What is your current place of residence?
North
3.0%
Central
12.1%
South
27.3%
West
36.4%
East
21.2%

8. In your opinion, what is the most significant benefit of incorporating AI into health insurance claims processing?



Count of 8. In your opinion, what is the most significant benefit of incorporating AI into health insurance claims processing?

Thematic Content Analysis

Role of AI In Analysis of Insurance Claims Based on Health

- **Theme 1: Improved Efficiency and Accuracy -** AI technologies enhance the speed and accuracy of claim analysis.
- **Theme 2: Fraud Detection and Prevention -** AI is crucial in identifying and preventing fraudulent insurance claims.
- **Theme 3: Predictive Modeling for Risk Assessment -** AI models assist in predicting health-related risks, influencing premium rates.
- **Theme 4: Enhanced Claim Verification -** AI-powered systems streamline the process of verifying claim information.

Enhancing Disease Risk Prediction with ANNs

- **Theme 1: Enhanced Disease Risk Prediction -** ANNs improve the accuracy of predicting disease risk based on multi-dimensional data.
- **Theme 2: Data Integration Challenges -** Integrating diverse data sources into ANN models is both a necessity and a challenge.
- **Theme 3: Feature Selection for Improved Models -** Selecting relevant features and variables is essential for optimizing ANN performance.
- **Theme 4: Interpretability vs. Accuracy -** A trade-off exists between model accuracy and the interpretability of results.

Impact of Lifestyle Factors In Claim Prediction

- **Theme 1: Lifestyle Factors and Risk Assessment** - Lifestyle factors play a crucial role in assessing the risk associated with insurance policyholders.
- **Theme 2: Premium Calculation and Lifestyle Choices** - Insurance companies factor in lifestyle choices when calculating premiums.
- **Theme 3: Claim Frequency and Lifestyle-Related Claims** - Lifestyle factors contribute to the frequency of certain types of claims.
- **Theme 4: Data Collection Challenges** - Gathering and utilizing lifestyle-related data presents challenges for insurers.

Geographic Variations in Healthcare Costs

- **Theme 1: Supply and Demand Dynamics** - Geographic differences in healthcare costs are influenced by variations in the supply of and demand for healthcare services.

- **Theme 2: Healthcare Infrastructure and Resource Allocation** - The availability of healthcare facilities and resources varies by location, impacting cost

- **Theme 3: Population Demographics and Health Needs** - Demographic factors and health needs differ across regions, affecting healthcare utilization and costs.

Demographics and Health Outcomes

- **Theme 1: Social Determinants of Health** - Demographics, such as socioeconomic status, education, and living conditions, significantly influence health outcomes.
- **Theme 2: Health Disparities** - Health disparities based on demographic factors contribute to differential health outcomes across populations.
- **Theme 3: Access to Healthcare Services** - Demographics play a crucial role in determining access to healthcare services and, consequently, health outcomes.

IMPLEMENTATION OF ANN(Artificial Neural Network)

DATASET LINK:
https://www.kaggle.com/datasets/thedevastator/insurance-claim-analysis-demographic-and-health

- As our topic is ‘HEALTH INSURANCE PREDICTION AND ANALYSIS BY ARTIFICIAL NEURAL NETWORKS’, in the coming few slides we will look at the practical application of ANN .

- **Application of ANN in health insurance prediction** : ANN is used to automate the process of claim approval by assessing the validity of claims based on historical data and predefined criteria. This can improve the efficiency of claims processing while reducing human error.

Colab link:
https://drive.google.com/file/d/1Xca2Ra7bOLcClseRBK2B4bmvbhXxReTN/view?usp=sharing

- Here are some important findings of our ANN model :

- The types of error in the model are,

```
Best Mean Squared Error: 16604623.733819183
mean_absolute_error: 1106.1495
```

The R2 score are:

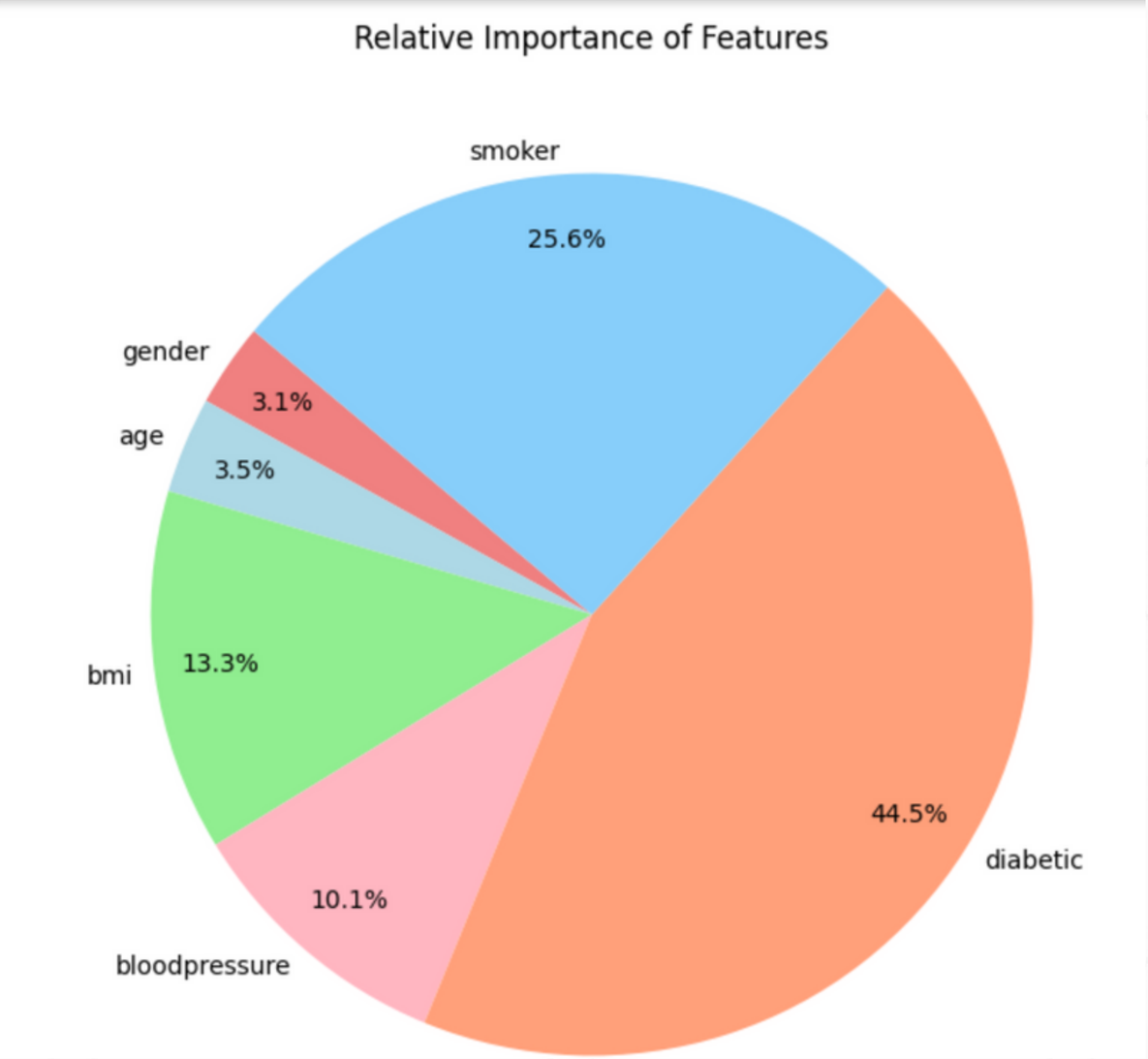
```
training R-squared Score: 0.9801734284272754
```

```
testing R-squared Score: 0.9831007292485577
```

```
Validation R-squared Score: 0.9700463202984729
```

Pie chart of importance of each feature.

As it is clearly seen, a person having diabetes and his/her smoking habits lead to increase in issuing of claims



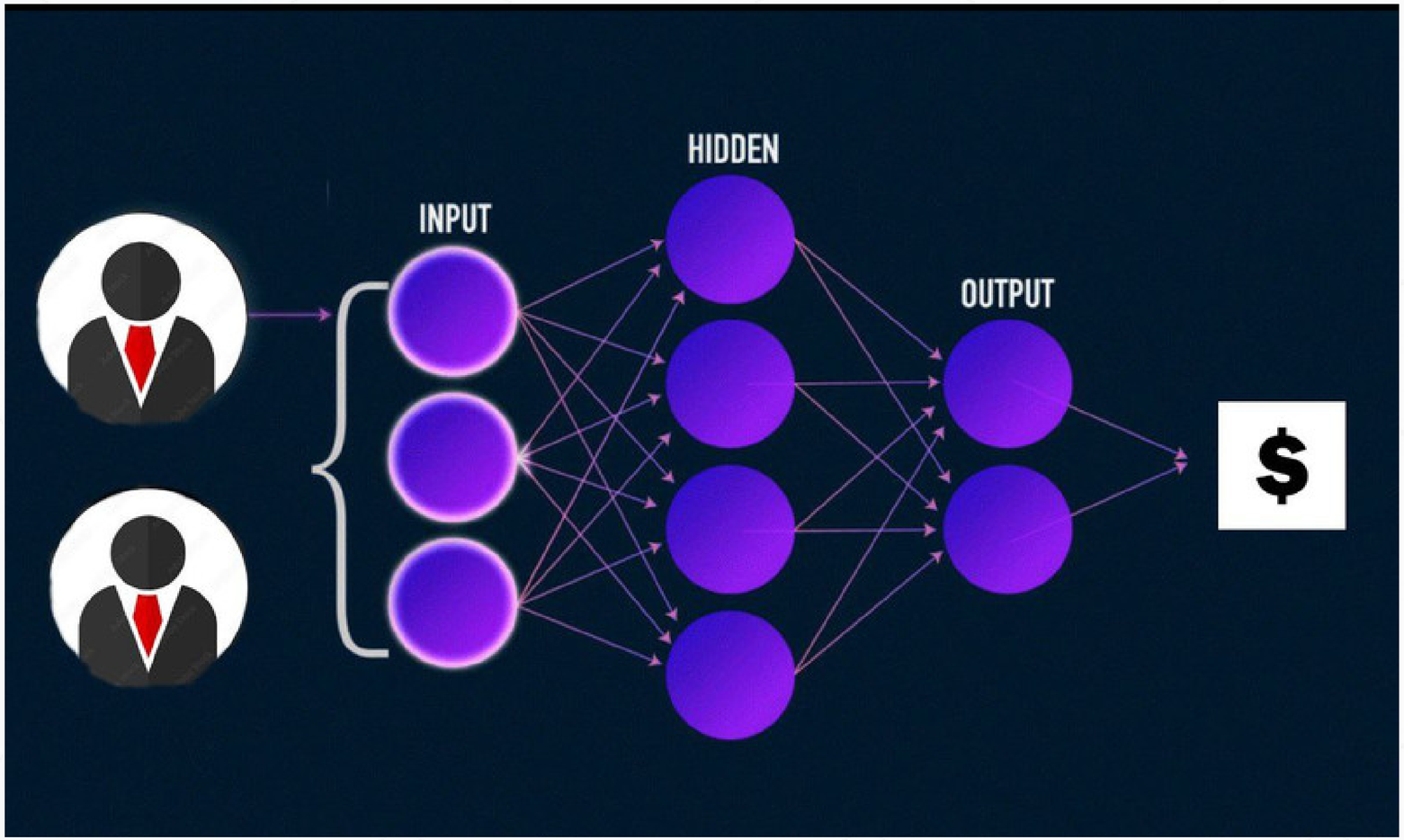
Basically Our ANN model looks like:

INPUTS

1)Age
2)BMI
3)Blood Pressure
4)Diabetes
5)Smoking

OUTPUT

Claim Charges



CONCLUSION

In our study, we aimed to apply ANN to health insurance claim forecasting and analysis, with the main focus on enhancing prediction accuracy and enabling informed decision-making.

MAIN CONCLUSIONS:

- Artificial Neural Networks (ANNs) have demonstrated their ability to uncover complex relationships and patterns, enabling highly accurate predictions of claim outcomes
- AI can considerably increase the accuracy of health insurance prediction
- Smoking and alcohol consumption might have an impact on a policyholder's health and thus the amount of a claim.
- Healthcare costs are influenced by geographical location.
- Based on different demographic groups (e.g., age, gender), there are variations in health insurance claims.

RESEARCH LIMITATIONS

- **Data Challenges:** The availability and quality of the data can vary, which can impact our findings.
- **Privacy and Security:** Restricting our use may be necessary to protect personal data.
- **Complex Models:** Although very effective, our prediction models can be difficult to understand.
- **Sample Size Matters:** The size of our data can influence how confident we are in our predictions.
- **Data Selection:** The data we used may not perfectly represent everyone.
- **External Influences:** Real-world events can affect our predictions.
- **Variable Relationships:** The interactions between factors can be quite complex.

FUTURE RESEARCH

- Investigate advanced ANN architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), for more accurate risk modeling.
- Develop real-time fraud detection systems using ANN models to quickly respond to fraudulent claims.
- Research on predicting customer churn and implementing strategies to retain customers.
- Develop chatbots using ANNs to improve customer service and automate responses to common inquiries.
- Research on using ANNs for assessing cybersecurity risks and predicting potential data breaches and cyberattacks.

THANK YOU