INDUSTRIAL AND SOCIAL PSYCHOLOGY BEHIND THE USE OF DATA SCIENCE AND MACHINE LEARNING IN THE FINANCE SECTOR.

Group Number - 1

Name	Reg. No.
Aakash Chaurasiya	20BCS001
Abhishek Dubey	20BCS002
Adithya Dileep	20BCS005
Adithya Hegde	20BCS006
Aditya Tiwari	20BCS007
Allu Hanuma Reddy	20BCS008
Gitesh Kailas Ambre	20BCS009
Vivek Yadav.P	20BCS105

Under the Guidance of,
Dr. Rajesh N S
Assistant Professor
Department of Humanities and Science
Indian Institute of Information Technology
Dharwad

November, 2021

Contents

Abstract

This report corresponds to the research conducted to analyse students' opinions on the use of these cutting edge technologies in banking services and focuses on aspects like ease of use, consumer privacy, security and transparency with respect to use of consumer data.

This report also includes the research conducted to review several articles and reports from reputed journals and publications which focus on the benefits that machine learning and data science bring to both business and consumers, which are reasons for their widespread use in this day and age. And we came up with some gaps where several studies and journals did not explore the opinions of young consumers on the use of data science and machine learning in banking and finance.

Here we came to know that introduction of artificial intelligence to the banking industry makes it possible to help customers more quickly and with lesser human intervention.

Here we've attempted to cover the customer response to the increased data science and machine learning use in the banking sector. And we've collected their opinion based on their experience in banking and other finance related sectors and analysed the data.

Introduction

Data science is the domain of study that deals with vast volumes of data using modern tools and techniques to find unseen patterns, derive meaningful information, and make business decisions.

Machine learning is the study of computer algorithms which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy of prediction by feeding on data and information in the form of observations and real-world interactions. Machine learning plays a major and important role in the field of data science. Data science uses complex machine learning algorithms to build predictive models.

The humongous amount of financial data often amounting to petabytes of records available due to digitization of customer data and financial services has made Data science and machine learning feasible since both require large amounts of data to provide reliable results.

There are many applications of data-science and machine learning in the banking sector and some of them include: Risk Modelling, Fraud Detection and Prediction of Inflation all of which are most critical to the functioning of essential banking services.

The increase in the usage of data science in recent years in the finance sector, especially in the banking sub sector has changed the banking industry in huge ways. The usage of data science in banks has now become an absolute necessity in order to be able to keep up with the competition.

With the growing number of banks, customers expect top notch operational efficiency from banks and hence, banks are looking at various better ways to understand customers and retain them.

The availability of transaction data, purchase history, brand loyalty and other data makes it ripe for the application of data science, and banks do so in order to allow them to get valuable insights into various customer related aspects. These insights further help the banks to make business decisions and to provide efficient services and new features to their customers. It also helps them in efficient resource utilisation and in engaging with customers more meaningfully.

Hence, data science must be involved in the banking sector, in their business and decision making processes to stay ahead of the competition and to ensure customer satisfaction and improve efficiency.

Research Questions

1. What is the value that data science and machine learning bring to businesses and consumers which makes them so widespread?
2. What is the opinion of young consumers about the use of data science and machine learning in the banking and finance sector?
3. What are the opinions of young consumers about the privacy and use of their data with respect to big data practices?
4. How aware are young consumers of the different applications of machine learning and data science in the banking and finance sector?

Research Objectives

The objective of this report is to research how data science is used in the banking sector, and its various applications and growing trends. Also to carry out a survey to find out the opinion of customers (students aged 18-25, to be more specific) about banking and digital banking, and also about their opinion about how banks use their data and the knowledge they have with respect to the technologies that use ML and data science. This will allow banks to model their technology to be more suited to their customer base in the coming years.

Literature Review

ML and Data Science applications in the finance and banking sector have been thriving in recent years.

Praveen Kumar Donepudi, (2017)[1] in his paper talks about how there's a paradigm shift in the world's financial institutions that can most notably be seen in Fintech organisations. He references the Q2 report of India's fintech industry to highlight the increase in investments, showing the growth of the financial sector. Traditional banking institutions are adopting computational intelligence technologies like AI chatbots, AML(anti-money laundering) etc. and it is also being employed in applications such as fraud pattern detection, personalized banking, automation, risk management and customer recommendation (Yu et al, 2016)[2]. He also references a research carried out by BCG consulting group that shows how China has applied AI and ML applications in its fintech sector and how it is significantly ahead of other countries. The research also predicts that 37% of the job market will have changed in this sector with ML and AI providing increased automation and efficiency. In short, these technologies are taking over the banking sector, with customers realizing that technology is not hard to learn or expensive, financial institutions have no choice but to catch up.

Sebastian Doerr & Leonardo Gambacorta & José María Serena Garralda, (2021)[3] mainly talk about how big data and machine learning is used in central banking, referencing a survey [4] carried out by the IFC(Irving Fischer Committee). They talk about the insights it offers: the nature of the definition of Big data these banks use, how the increase in interest about big data has led to its implementation in various projects and is used in economic research and policy decisions. The survey also uncovers the various obstacles to big data usage such as cleaning raw data, sampling, ethics, privacy issues, budget constraints etc. The authors talk about how traditional data is preferred to be complemented with non-traditional data sources. Through the survey they talk about the various purposes for which big data is used by central banks which include economic research, monetary policy, financial stability analysis, suptech/regtech and other uses. There are however, challenges like creating an IT infrastructure, ethical and privacy

concerns that banks need to look into. In regard to usage of big data for predictions, the paper highlights the trade-off between accuracy and explainability, and how interpretable Ml algorithms are working on this shortcoming. Another interesting point of note is how banks recognise the role of policy cooperation, and are willing to share data based on expertise, to use big data to work on global issues and to develop exploratory projects and share financial and human resources.

Bholat, David, Mohammed Gharbawi, and Oliver Thew.(2020)[5] conducted a survey to understand the impact of the Covid-19 pandemic on the use of machine learning and data science on the banks in the UK. The survey reports that the use of and interest in these computation intelligence technologies has remained mostly stable, and in some cases has actually increased. Most banks are certain that there will be an increase in investment in data science in the coming years. In the pandemic, many banks carried out measures to provide financial stability to businesses through guaranteed loan applications which is beneficial to bak operations. The banking sector has not suffered from the pandemic to the extent of other industries. Some banks have reported a negative impact on ML model performance, which might mostly be due to problems with recalibrating these models. However, remote working has allowed businesses to acquire more data which has led to an increase in the demand for data scientists and engineers.

Research Gap

A gap in research that we've attempted to cover is the customer response to increased data science and machine learning use in the banking sector. There hasn't been any literature reviewing the opinions of the age group who will be the largest customer base in coming years. Machine learning and data science applications are expected to find increased use in the coming years, so the opinions of the future customers is quite significant.

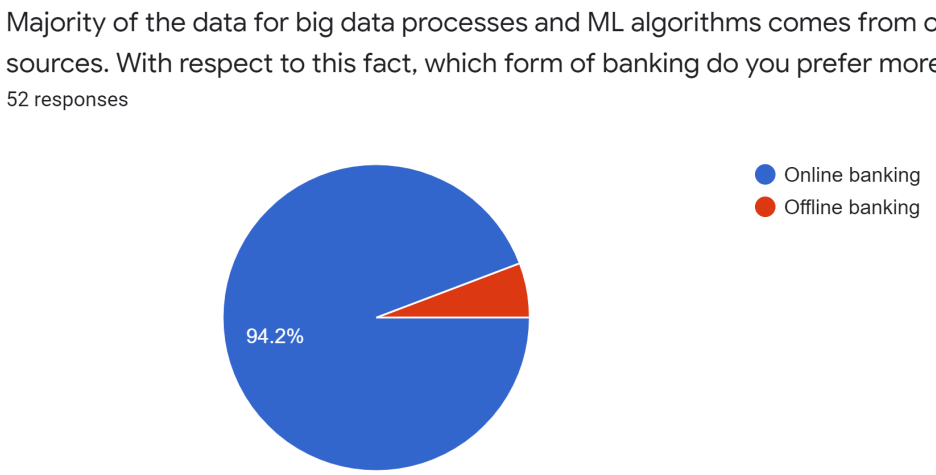
Research Methodology

For research we reviewed some articles and reports from reputed journals and publications which focus on the many benefits that machine learning and data science bring for both businesses and consumers. We also conducted a survey for our research which was sent to students through email and other social media channels. The survey was a Google form which was answered by a total of 52 students.

The questions were framed and chosen in such a way as to attempt to find a student's opinion on aspects relating to ease of use, consumer privacy, security and transparency with respect to use of consumer data. The responses of the students were visualised by Google form's data visualization software in the form of pie charts and bar graphs, providing insights into the responses of the students.

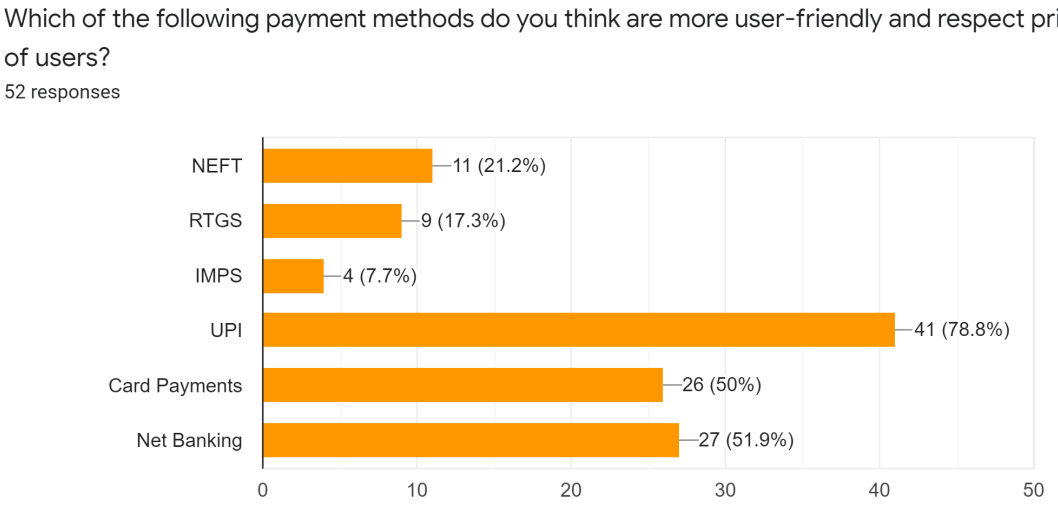
Results and Analysis

Question 1



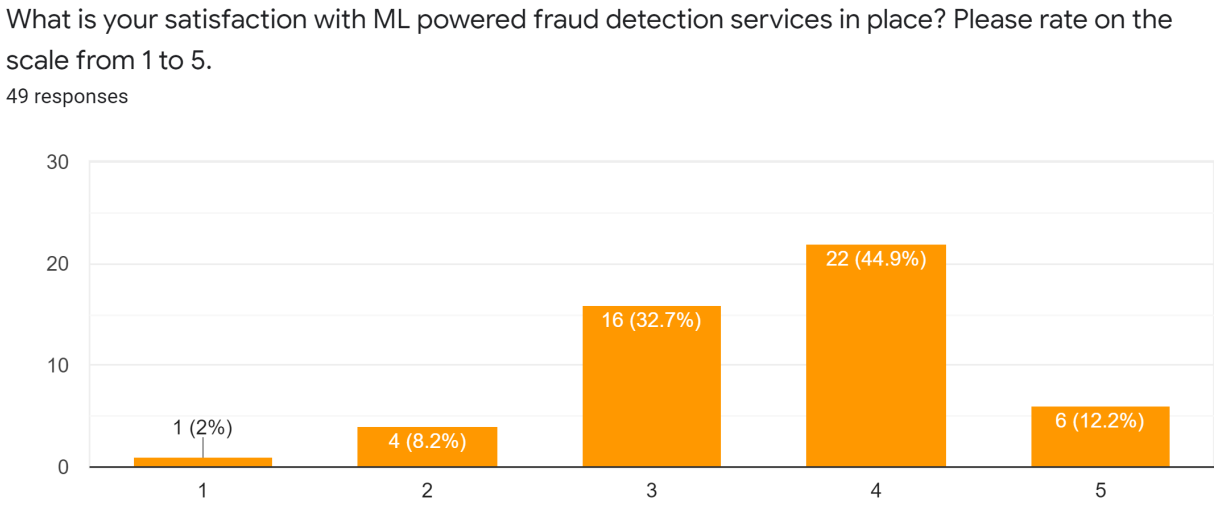
94.2% of respondents preferred online banking. It indicates that most people don't want to do the offline transactions which is good news for a business organisation, since to make good predictions we need to collect data from customers and offline it is very difficult because the written data on forms and documents has to be digitized either manually or with the help of electronics for the purpose of data science.

Question 2



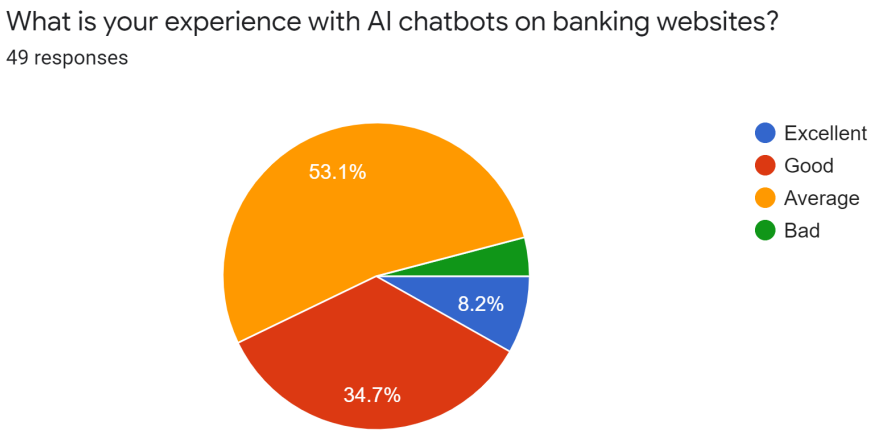
Even though NEFT, RTGS, and IMPS allow transfer of large sums of money and provide more security relative to most services, users opt for the more user-friendly, and hassle-free payment methods like UPI and card payments. Also, newer systems like net banking which have evolved from the old methods of payment are seen to be more preferable.

Question 3



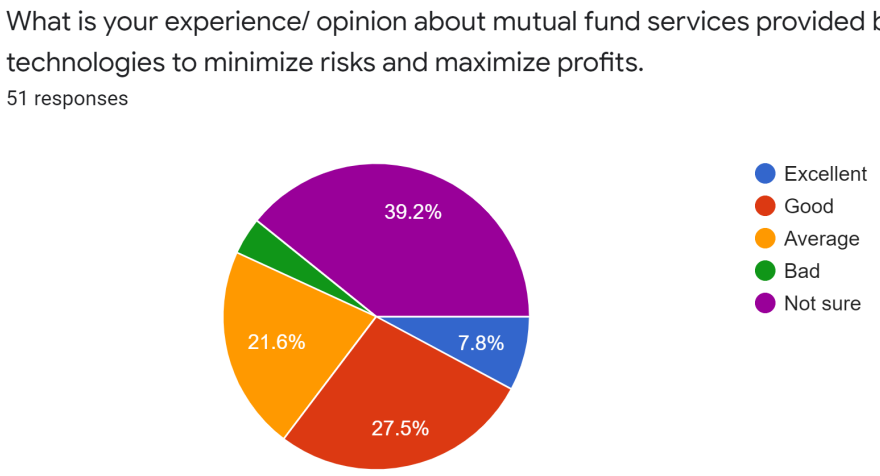
More than 75% of the users have given a rating of 3 or above. Despite a large number of online frauds that have been seen in recent times, the results show that students, on average, are more than satisfied with the fraud detection systems in place.

Question 4



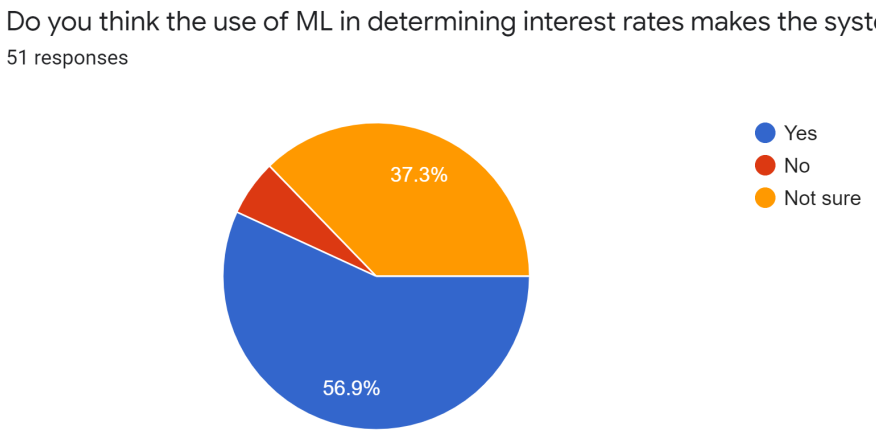
Most of the people, about 53.1% of people have an average experience, 34.7% of them have a good experience and 8.2% have an excellent experience while a very few, about 8.2% had a bad experience on using AI chatbots on banking websites. Actually, AI chatbots are different for each bank, and the experience of the customers depends on the banks. From this, it can be concluded that generally, the customers have an average to good experience with AI chatbots.

Question 5



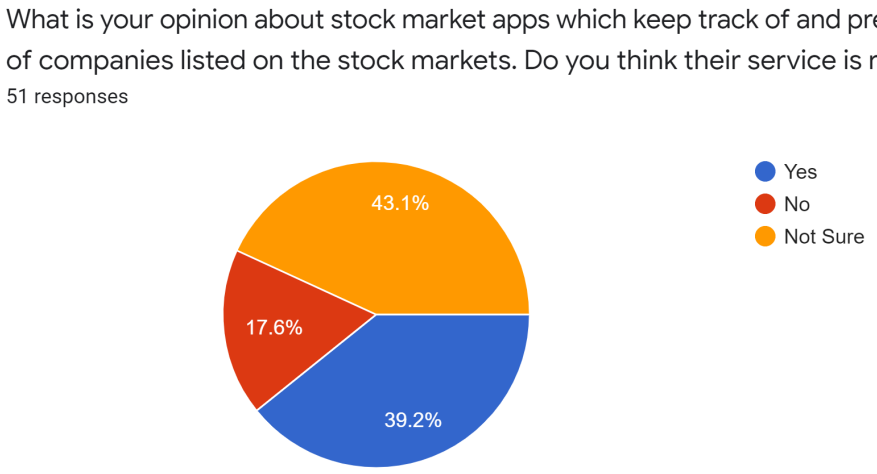
Most of the young consumers (about 39.2%) are not sure about it. It might be possible that they do not use mutual fund services provided by banks . And another 27.5% of them have a good experience and 21.6% of them have an average experience and only 7.8% have had an excellent experience. Therefore the collective opinion of young consumers about mutual fund services is a mixed bag.

Question 6



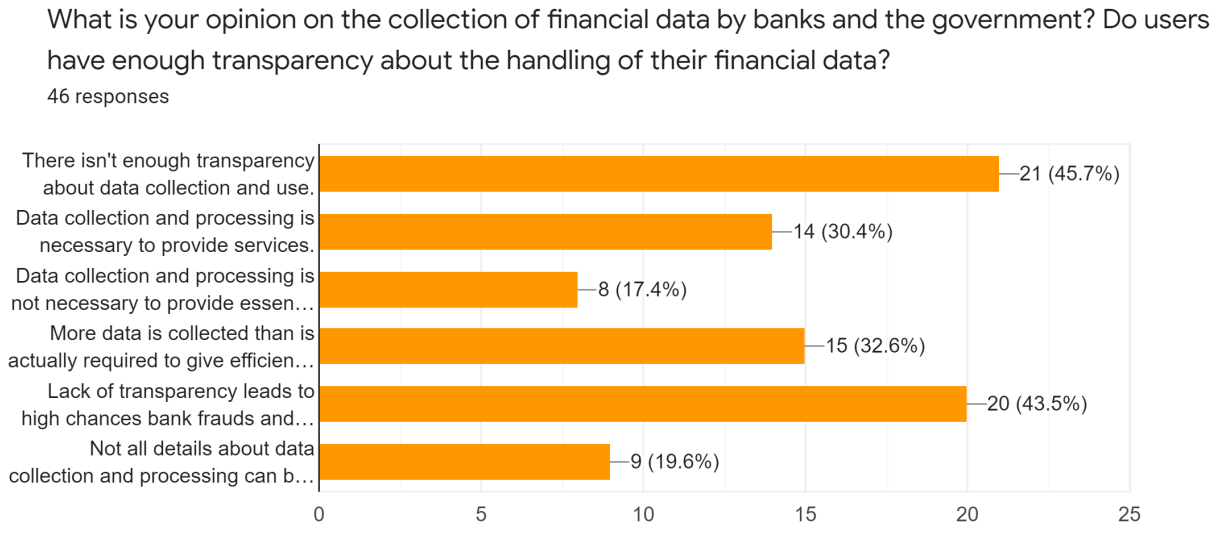
57% of the respondents feel that the use of machine learning in determining interest rates makes the banking system efficient while approximately 6 percent respondents don't. One third of the respondents aren't sure about whether machine learning should be used for determining interest rates. This shows that while the majority of young consumers approve of the use of machine learning, a significant fraction either doesn't approve or is not sure. Awareness of the use of machine learning in prediction of crucial market parameters like interest and inflation rates is increasing and we might see more approval towards it in the future.

Question 7



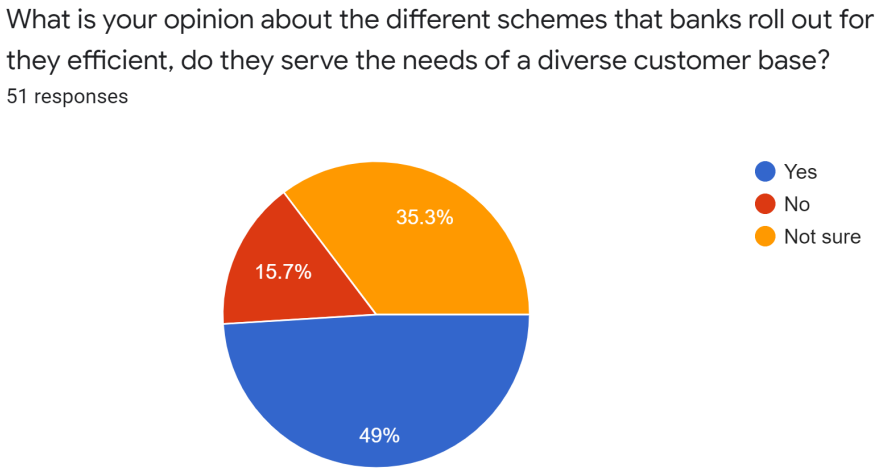
Almost 50% of the users were seen to not have enough knowledge about the stock market apps that kept track of stock prices and were hence not sure about the reliability of these apps. A significant number of the rest however have also found these apps unreliable. Algo trading softwares are being developed and improved constantly and perhaps will prove to be more reliable in the future.

Question 8



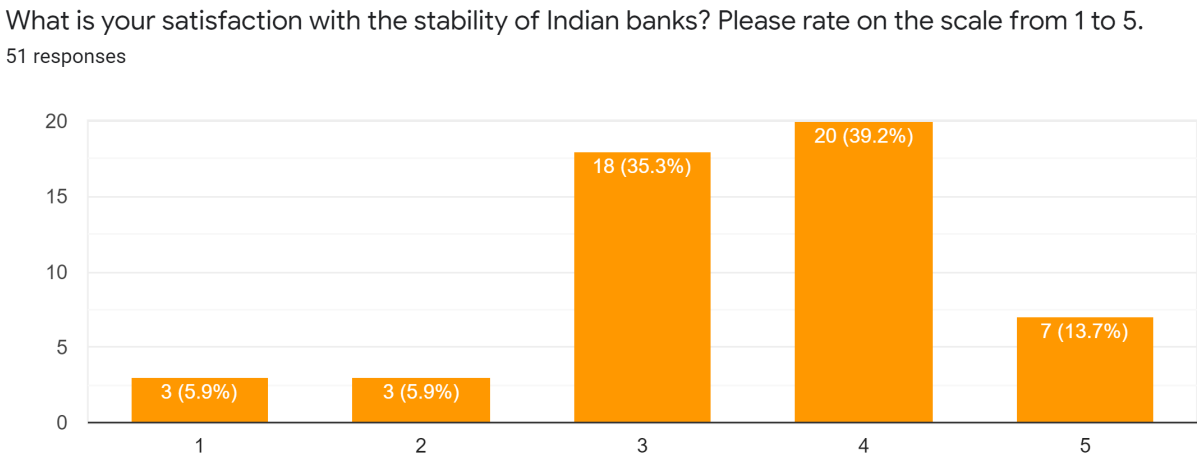
When asked about the opinions about financial data collection and the transparency offered in regard to that data, the responses obtained were seen to be quite varied. A large number of respondents felt that there wasn't enough transparency about data collection and use and that the lack of transparency can lead to high chances of bank frauds and unethical government surveillance. Quite a number of users agreed that data collection and processing is necessary to provide services, however more data is being collected than necessary to provide essential services. There were some who responded that data collection and processing is not necessary to provide essential services. Consumers were hence seen to be okay with their data being used to provide efficient services, and some even understand that not all details about data collection and processing can be disclosed for the security of bank and government services.

Question 9



The responses that we got from the above question indicates that the needs of most of the customers are served by different schemes that the banks are rolling out. It shows that with the help of predictions made about the needs of customers using data science and machine learning algorithms, the banks are able to make better decisions that help their customers from diverse socio-economic backgrounds and provide better services to them. Also improvements can be made since 15.7% of the respondents were dissatisfied with the efficiency of the services provided by banks. One third of the respondents were not sure, meaning they might not be aware of the various schemes rolled out by banks for people from diverse backgrounds.

Question 10



Majority of the respondents have given the rating of 3 and above. This shows that almost 90% of the people are satisfied with the stability and working of Indian banks. But there also seems to be scope for improvement as the number of respondents choosing the rating of 5 is very limited and on the other hand, more than 10% of youth consumers are not satisfied with the stability of Indian banks. Therefore the performance of the Indian banking system is on a better side with some scope for improvement.

Conclusion and Summary

Data Science and machine learning applications in the financial sector and the banking industry have been rising significantly in recent years, and will only continue to do so. Companies have no option but to adapt data science models in their business and decision making processes to stay ahead of the competition and to ensure customer satisfaction and improve efficiency. In this research we've attempted to find the opinion of the future customer base (i.e on students aged 18-25) on aspects relating to ease of use, consumer privacy, security and transparency with respect to use of consumer data. The findings show that a large number of users are satisfied with the application of machine learning and data science in digital banking. However in the cases of stock markets or mutual funds, many users are not very sure about the way in which their data is used. They understand that banks have to collect data and the details about this collection and processing can't always be revealed because of security concerns, but feel that there could still be increased transparency. As the usage of data science will continue to rise rapidly it becomes absolutely necessary for banks to pay more heed to the concerns and opinions of the younger consumers, since they will serve as the majority customer base in the coming years.

References

1. Donepudi, Praveen Kumar. "Machine learning and artificial intelligence in banking." View of Machine Learning and Artificial Intelligence in Banking. *Engineering International* 5.2 (2017): 83-86.
2. Yu, L., Yang, Z., & Tang, L.(2016). A novel multistage deep belief network based extreme learning machine ensemble learning paradigm for credit risk assessment. *Flexible Services and Manufacturing Journal*, 28(4), 576-592.
3. Sebastian Doerr & Leonardo Gambacorta & José María Serena Garralda, 2021. "Big data and machine learning in central banking," *BIS Working Papers 930, Bank for International Settlements.*
4. IFC Report No 13, Use of big data sources and applications at central banks 2020 survey conducted by the Irving Fisher Committee on Central Bank Statistics (IFC). *Bank for International Settlements.*
5. Bholat, David, Mohammed Gharbawi, and Oliver Thew. "The impact of Covid on machine learning and data science in UK banking." *Bank of England Quarterly Bulletin* (2020): Q4.

Annexure

To get an idea of the opinions of the customer about banking and digital banking we conducted the survey of students in the age range of 18-25. The following questions were sent to students.

1. Majority of the data for big data processes and ML algorithms comes from online and digital sources. With respect to this fact, which form of banking do you prefer more?
    a. Online Banking
    b. Offline Banking

2. Which of the following payment methods do you think are more user-friendly and respect the privacy of users?
    a. NEFT
    b. RTGS
    c. IMPS
    d. UPI
    e. Card Payments
    f. Net Banking

3. What is your satisfaction with ML powered fraud detection services in place? Please rate on the scale from 1 to 5.

4. What is your experience with AI chatbots on banking websites?
    a. Excellent
    b. Good
    c. Average
    d. Bad

5. What is your experience/ opinion about mutual fund services provided by Banks which use ML technologies to minimize risks and maximize profits.
    a. Excellent
    b. Good
    c. Average
    d. Bad
    e. Not sure

6. Do you think the use of ML in determining interest rates makes the system efficient?
    a. Yes
    b. No
    c. Not Sure

7. What is your opinion about stock market apps which keep track of and predict the prices of stocks of companies listed on the stock markets. Do you think their service is reliable?
    a. Yes
    b. No
    c. Not Sure

8. What is your opinion on the collection of financial data by banks and the government? Do users have enough transparency about the handling of their financial data?
    a. There isn't enough transparency about data collection and use.
    b. Data collection and processing is necessary to provide services.
    c. Data collection and processing is not necessary to provide essential services.
    d. More data is collected than is actually required to give efficient services.
    e. Lack of transparency leads to high chances of bank frauds and unethical government surveillance.
    f. Not all details about data collection and processing can be disclosed for the security of bank and government services.

9. What is your opinion about the different schemes that banks roll out for different customers, are they efficient, do they serve the needs of a diverse customer base?
    a. Yes
    b. No
    c. Not Sure

10. What is your satisfaction with the stability of Indian banks? Please rate on the scale from 1 to 5.

Survey link: https://forms.gle/ut9Dk9m6SVrSxonC9