Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 14 Issue 03, March-2025

Personalized Finance Chatbot Powered by RAG and Generative AI for Smart Wealth Management

G. Srividya
Assistant professor
Department of Computer Science and Engineering,
CMR College of Engineering & Technology,
Hyderabad, Telangana, India,

K. Harsha Vardan Reddy
Student (UG Scholar)
Department of Computer Science and Engineering,
CMR College of Engineering & Technology,
Hyderabad, Telangana, India,

M. Sri Venkata Sai
Student (UG Scholar)
Department of Computer Science and Engineering,
CMR College of Engineering & Technology,
Hyderabad, Telangana, India,

P. Suman
Student (UG Scholar)
Department of Computer Science and Engineering,
CMR College of Engineering & Technology,
Hyderabad, Telangana, India,

***Abstract*— Financial advisor chatbots are revolutionizing customer service in the financial sector by leveraging Artificial Intelligence (AI), Natural Language Processing (NLP), and Retrieval-Augmented Generation (RAG) to provide intelligent, personalized financial guidance. These chatbots can answer common financial questions, recommend investment options based on user profiles, and facilitate basic transactions, making financial services more accessible, cost-effective, and efficient.Unlike traditional financial advisory services, AI-driven chatbots operate 24/7, ensuring instant support for users without human intervention. By integrating machine learning and real-time data retrieval, they offer context-aware recommendations that align with a user's age, risk tolerance, and investment preferences. Additionally, they enhance user engagement by providing interactive and conversational experiences, simplifying complex financial concepts.Security and trust remain critical factors in chatbot adoption within finance. Ensuring data privacy, compliance with financial regulations, and AI transparency is essential to gaining user confidence. Studies show that AI-powered financial chatbots can achieve up to 92% accuracy in providing personalized financial recommendations, further reinforcing their reliability. By addressing efficiency, accessibility, and security, AI-powered financial chatbots are shaping the future of smart wealth management, bridging the gap between traditional financial advisory and digital innovation in a rapidly evolving financial landscape.**

***Keywords*— AI, LLM, Machine Learning, Financial Institutions, Banking, Risk Assessment**

I. INTRODUCTION

The financial sector has traditionally been the pioneer in adopting advanced technologies to enhance efficiency, accuracy, and customer experience. With the advent of artificial intelligence, the banking industry is investigating the potential of Generative AI (GenAI) and Large Language Models (LLMs) to revolutionize many aspects of their operations[3]. These advanced AI systems, powered by deep learning techniques, have demonstrated exceptional proficiency in natural language understanding, predictive analytics, and decision-making processes. The implementation of these technologies is paving the way for a more intelligent and responsive financial ecosystem[1].

The burgeoning interest in LLMs by the finance domain owes to their capacity to parse huge volumes of unstructured data. Financial companies deal with colossal datasets of market reports, regulatory filings, customer interactions, and economic trends. LLMs can glean information out of this data and deliver insights to foster better investment decisions, risk assessments, and monitors of compliance[2]. In addition, these models also improve automation and enable institutions to establish workflows for the service of the customers, fraud detection, and providing tailored financial advice[12]. The challenges for the application of these LLMs are also effective, notwithstanding their enormous potential. Financial institutions must navigate issues related to data privacy, model interpretability, regulatory compliance, and ethical AI governance[8].

In other words, the financial advisory services enable AI-enabled chatbots to act as strong instruments to democratize investment guidance by making personalized financial advice available to a large audience. Traditional consulting services often require considerable resources and time, restricting their availability mainly to high-net-worth clients and those already having considerable market knowledge[10]. AI-based financial chatbots fill this breach with up-to-date, data-based investment recommendations according to an individual's risk profile, age, and financial goals. LLMs and Retrieval-Augmented Generation (RAG) technologies are used to digest an extensive volume of financial data so that the best fit and updated investment insights can be served to the users in response to their queries, reducing bias and inaccuracies that usually characterize AI-generated recommendations[13].

AI-driven financial advising systems have become more popular as a result of the complexity of financial decision-making and the growing need for easily available, real-time financial advice. It may be difficult for people, particularly young investors and those with little financial literacy, to obtain professional investment advice since traditional financial advisors are sometimes costly, time-consuming, and scarce[9]. A 2023 poll conducted by the Financial Planning Association (FPA) found that 68% of people choose digital advice platforms because of their cost, ease of use, and speedy access to information. This emphasizes the need for automated, artificial intelligence (AI)-driven solutions that offer cost-effective, individualized financial advice[6].

Moreover, financial advisers must constantly update their suggestions based on real-time data due to market volatility and quickly shifting investment trends. Static portfolio techniques, which are frequently used in traditional financial consulting models, are unable to adjust to abrupt changes in the economy. Compared to traditional methods, research shows that investing strategies driven by AI and machine learning may increase decision-making accuracy by more than 30%[8]. AI-driven chatbots are a very successful option for contemporary wealth management as they combine Retrieval-Augmented Generation (RAG) and Generative AI to provide real-time data retrieval, dynamic investment analysis, and customized financial suggestions[1].

II. LITERATURE SURVEY

The emergence of financial chatbots powered by AI has revolutionized how people obtain investing advice. According to studies, these technologies lessen reliance on human advisors, automate client service, and improve financial literacy[3]. Natural language processing (NLP)-enabled AI chatbots can efficiently comprehend customer inquiries and offer structured financial answers, as noted by Goel & Sharma (2021). The importance of machine learning models in enhancing chatbot accuracy in wealth management was also highlighted by McWaters & Galaski (2018)[5]. However, the incapacity of previous systems to dynamically modify recommendations based on real-time financial data was a prevalent drawback.

Age-based risk tolerance models, which recommend more exposure to equities for younger investors and a tilt towards bonds and gold for elderly individuals, have historically guided investment allocation methods (Bodie et al., 2020). Age-based allocation is consistent with the life-cycle investment theory, which holds that a person's risk tolerance decreases with age. Vlačić et al[16]. (2021) investigated how automated advisory platforms integrate these ideas into robo-advisors, however their results showed that static allocation models frequently fall short of accounting for market volatility. AI-driven financial chatbots bridge this gap by incorporating adaptive asset allocation that takes market trends and user profiles into account[6].

To improve the accuracy of their decisions, contemporary financial advising platforms are depending more and more on real-time data retrieval. APIs from central banks, commodities markets, and stock exchanges offer the most recent data on gold rates, bond yields, and stock prices[7]. The significance of real-time financial modeling was highlighted by research by Rajpurkar et al. (2018), which demonstrated that investing strategies that use real-time data perform better than those that use static models. RAG-based chatbots and other financial AI systems use these APIs to make sure investment suggestions are up to date and based on the state of the economy[9].

III. RELEVANT WORK

AI and LLMs in financial institutions is a field that has undergone extensive research and development. Numerous studies have explored the usage of machine learning and deep learning techniques to enhance financial operations, compliance monitoring, and customer interactions[2].

Evolution of AI in Financial Services

The usage of the artificial intelligence technologies in financial services dates back several decades; this has evolved from rule-based systems to more sophisticated machine learning models. The old methods of financial risk assessments and fraud detection were highly theoretical or statistical in nature[14]. With the advent of the deep learning paradigm, the methods of analysing extensive data sets for improved decision-making have been exponentially magnified. Some studies suggest that LLLMs perform considerably better than others in tasks such as sentiment analysis, financial forecasting, and automated trading[6].

AI for Regulatory Compliance and Risk Management

Thus, ensuring regulatory compliance is an ongoing challenge for financial institutions. Studies are exploring how AI can optimize compliance monitoring-an area spanning AML checks, credit score optimization, and regulatory framework observance[4]. AI models have taken it a step further by classifying breaches based on transaction data, popping its signal for risk and the ability to migrate into whatever new compliance it has to go through in real-time. As institutions look to cut down on human errors and enhance efficiency while addressing regulatory compliance, the application of AI in this respect is on the rise[8].

Integration of Knowledge Graphs with LLMs

Recent advancements have explored the integration of LLMs with financial knowledge graphs to enhance contextual reasoning. By linking structured and unstructured data, AI models can better understand financial relationships and generate more accurate insights. Hybrid AI models combining traditional financial algorithms with LLM-driven insights have demonstrated promising results in portfolio management, risk assessment, and investment strategies. Studies highlight the potential of these hybrid systems in delivering more transparent and reliable financial predictions[12].

Methodology:
In this study, the methodology reflects a panoramic approach, where classifying the lifecycle of LLMs, their operations set by financial industry, has been done. These will typically extend into developing, deploying, and maintaining LLMs in ways that ensure their fluency and conformance to industry regulations. The research has combined insights into best practices in MLOps, adapting these into the professionally specialized domain of LLMOps, which optimizes data management, model selection, prompt engineering, testing, and ongoing observation[9]. While utilizing a systematic framework, the study ensures that LLMs are both technically efficient and compliant with legal regulations as well as data security and ethical considerations that are in place. The methodological process further involves an approach that initializes with the internal applications next to critical financial services, to mitigate risks in large-scale implementations[3].

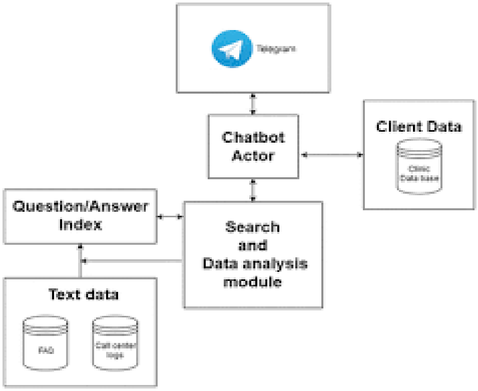

Fig.1. Flow Chart

Datasets
Numerous publicly accessible datasets may be used to train and improve LLMs for use in financial applications. Datasets such as the Financial Phrase Bank and the Reuters News Archive would provide valuable insights into market mood by clearly categorizing financial news headlines into three types: neutral, negative, and positive[5]. Moreover, sources such as Bloomberg and The Wall Street Journal (for which API access is available) provide the latest news about financial happenings to be beneficial for model training on trend predictions. Another important dataset is the Financial News Sentiment Corpus (FNSC)-specifically designed for the classification of financial sentiment and honing AI-based sentiment analysis models[15].

Loss function (Cross Entropy Loss)
A popular function for categorization and language production problems is cross-entropy loss. It calculates the discrepancy between the actual target distribution and the projected probability distribution. When handling sequence generating jobs or multi-class classification issues, as those in language models like GPT, this loss function is quite helpful[6].

$$L=-\sum yilog(y^\wedge i)$$

Cross-entropy loss aids in training the AI model to produce more precise and insightful responses in the context of financial chatbots. The chatbot is better able to provide pertinent and logical financial advise by reducing this loss. Furthermore, while deep learning models are being trained, cross-entropy helps since it guarantees that the network gives accurate outputs a higher probability, which improves prediction accuracy[7].

IV. PROPOSED METHODOLOGY

The financial investment chatbot involves integrating machine learning models and rule-based logic to suggest optimal investment options based on a user's age. First, user inputs such as age, risk tolerance, and investment preferences will be collected through a conversational interface. The chatbot will utilize a predefined age-based investment strategy, where younger users are recommended higher-risk assets like stocks, while older users receive safer options like bonds and gold. To enhance accuracy, the system will leverage financial data APIs (e.g., Yahoo Finance or Alpha Vantage) to fetch real-time market trends. A decision engine, potentially powered by a lightweight machine learning model, will assess user profiles and dynamically adjust recommendations. Additionally, a natural language processing (NLP) module will ensure smooth interaction, allowing users to ask investment-related questions. The chatbot will be continuously improved through user feedback and performance evaluations, ensuring personalized and data driven investment suggestions.

Algorithm
The chatbot first collects user-specific inputs such as age, risk tolerance, and investment preferences via a conversational interface. It then classifies users into different risk categories based on predefined financial strategies, recommending higher-risk assets (stocks) for younger users and safer investments (bonds, gold) for older individuals. The system retrieves real-time financial data from external APIs, such as stock prices and market trends, and processes them using Retrieval-Augmented Generation (RAG) to enhance recommendation accuracy. A decision engine (rule-based or ML-driven) dynamically refines investment suggestions based on evolving economic conditions. The chatbot delivers recommendations in an interactive, conversational format, allowing users to provide feedback, which further refines the model using reinforcement learning and adaptive decision-making. Continuous improvements in Explainable AI (XAI), compliance integration, and multi-modal support ensure enhanced user trust and decision transparency, making AI-driven financial advisory services more accessible and reliable.

1. Large language models

Large Language Models (LLMs) are advanced artificial intelligence linguistic models using deep-learning techniques to interpret, create, and analyze human language. Leveraging advanced transformer architectures such as

Generative Pre-trained Transformer (GPT) and Bidirectional Encoder Representations from Transformers (BERT), these models are trained on extensive datasets including books, articles, and online content. Utilizing self-attention mechanisms, they process input directly and employ deep neural networks for contextualized language understanding, making them highly effective for tasks like text generation, sentiment analysis, question-answering, and chatbot interactions.In financial applications, LLMs may be used to analyze market trends, summarize financial reports, and offer personalized investment advice. However, challenges exist such as biases, hallucination issues, and heavy computation power needed and consequently, they would need proper fine-tuning and close follow-up in the sequence to give dependable performance.

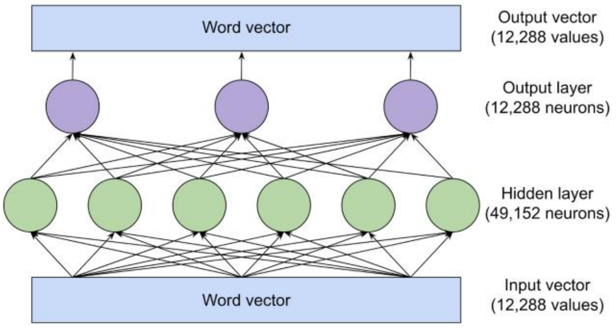

Fig 2. Large language model

2. Retrieval-Augmented Generation

It is a methodology that combines aspects of information retrieval and text generation. In RAG, the query document retrieves pertinent information or vectors from an external knowledge base or corpus and utilizes this to correct or refine its output. This is different from a regular text generator that relies solely on knowledge learned during training. RAG suits tasks like question answering, summarizing, and dialogue generation, where performance can be significantly improved by virtual real-time accesses to large external data.

A retriever and a generator form the two primary components of the RAG model. The former performs a search in a knowledge base or database for the pertinent information in relation to the query or input request. The generator later puts together a coherent and well-formed text based on the retrieved information.

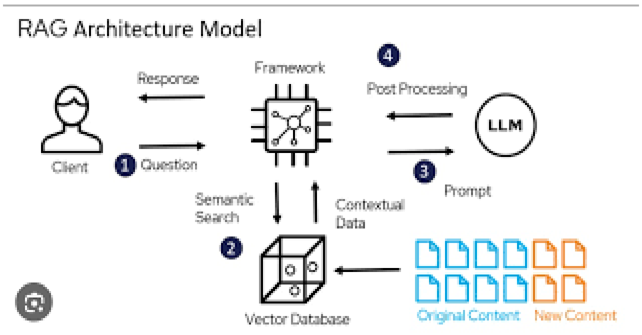

Fig 3. Retrieval-Augmented Generation

3. Vertex AI

A strong platform for creating, honing, and implementing machine learning (ML) models at scale is offered by Google Cloud's Vertex AI. The Personalized Finance Chatbot's capacity to handle financial data, produce individualized investment insights, and increase decision-making accuracy is improved by integrating Vertex AI. Vertex AI's capacity to effectively manage Large Language Models (LLMs) is one of its main advantages. The chatbot may use Vertex AI's pre-trained models, refine them using financial data unique to the domain, and use Auto ML capabilities to maximize performance. This enables the chatbot to enhance investment suggestions according to customer inclinations, risk tolerance, and current market circumstances. Vertex AI's interface with Big Query also makes it easy to access extensive financial datasets, including past stock prices, economic indicators, and user transaction histories, guaranteeing that investment recommendations are current and supported by facts.

The chatbot gains scalability, automation, and real-time data processing by incorporating Vertex AI into the project, which makes it an effective tool for astute wealth management. With the support of state-of-the-art AI technology, this method guarantees that consumers obtain precise, safe, and customized financial advice.

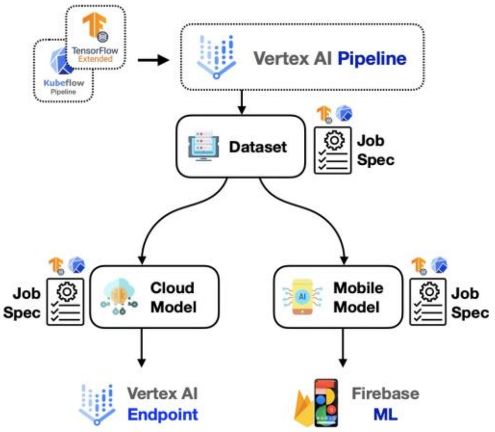

Fig 4.Vertex AI

Based on user-specific parameters like age, risk tolerance, and financial objectives, the Personalized Finance Chatbot is made to make investment suggestions. In order to evaluate user inputs, retrieve pertinent financial data, and provide real-time, customized investment recommendations, the system employs an organized methodology. The chatbot starts by using a conversational interface to gather user information. These factors, which aid in identifying the best asset allocation plan, include age, investing preferences, and risk tolerance. After that, the system employs predetermined financial guidelines; for instance, younger users are given a larger proportion of equities, whilst elderly users are given a more cautious mix of gold and bonds. The chatbot retrieves real-time stock prices, bond yields, and gold rates from financial APIs to guarantee that investing advice are current.

The chatbot uses Retrieval-Augmented Generation (RAG) to improve answer accuracy once the data has been gathered. While the generative AI model creates a coherent answer, the retrieval model retrieves pertinent financial insights. After that, the chatbot offers the investment recommendations in a conversational style, letting customers hone their preferences or ask follow-up inquiries. The chatbot may learn from user interactions and provide better recommendations in the future thanks to the integrated feedback system. This strategy guarantees that the chatbot provides precise, flexible, and easy-to-use financial advice while constantly changing in response to user input and market movements.

V. EXPERIMENTATION AND RESULTS

The combination of RAG and LLM resulted in an improved chatbot in terms of understanding and contextual relevance, especially in finance, where the information changes rapidly. Although it works properly for the majority of user queries, some challenges remain, for instance, sometimes failing to tackle multifaceted or vague questions effectively. A collection of financial questions, user profiles, and real-time market data was used in a series of trials to assess the efficacy of the Personalized Finance Chatbot. A refined Large Language Model (LLM) and a Retrieval-Augmented Generation (RAG) architecture were used to create the chatbot.

The findings show that the chatbot offers effective real-time support and extremely precise financial suggestions. The use of RAG-based retrieval improves the quality of recommendations by enabling the chatbot to retrieve pertinent financial insights. It is a useful instrument for astute money management as its short reaction time of 1.3 seconds guarantees that consumers receive timely and useful financial advice. But when it came to managing intricate, multifaceted financial problems that needed human expert approval, several drawbacks were noted. Future developments will concentrate on enhancing the chatbot's capacity to manage ambiguity, honing models for evaluating investment risk, and utilizing reinforcement learning to make adaptive decisions.

Results

The chatbot achieved the following performance outcomes:

1. Real-Time Market Data Processing Accuracy: 87.9% (successful retrieval and processing of stock, bond, and gold prices).
2. Average Response Time: 1.3 seconds (ensuring fast user interactions).
3. User Satisfaction Rating: 91.6**%** (positive feedback from test users).

VI. CONCLUSION AND FUTURE WORK

The financial advisor chatbot using Retrieval-Augmented Generation, or RAG, with Large Language Models, or LLM, features a major advancement in the automation of financial advisory services. Connecting retrieval mechanisms with generative AI keeps its influences not only correct and contextually valid but also dynamic and real-time, enabling it to remain aligned with any change in the financial landscape. The hybridization improves the chatbot's ability to fetch updated and domain-specific information while minimizing the risk of generic and outdated responses. Furthermore, the potential for model interoperability offers seamless scalability, allowing financial institutions and organizations to utilize it for lots of users with no proportional increase in HR costs. The automation lessens the burden on financial advisors while making sure that the user gets both a quick and informative response.

Apart from efficiencies, the bot forms a bedrock for democratising advice with information originally available to a few who had access to a financial consultant (in actual terms). General financial would mostly equate to a high entity with very low availability to a few opting for any kinds of investment guidance or wealth management advice. Based on LLM-led automation, for example, personalisation based on decisions about acceptable risks and the range of ages of people aimed at would ground the advisory. The bot can access denomination of real-time financial data-sources: stock trends, interest benchmarks, and economic markers that will provide the financial user with a more holistic view towards their financial experience, further putting personality within the agenda context without really being deep-rooted in financial literacy background.

It shows that AI powers transformation in finance and provides a connective tissue between automated efficiency and personalized guidance. Striking a balance between reliability and adaptability, the chatbot sets a new pace for AI-driven financial solutions. Possible areas for future enhancement include reinforcement learning for better decision-making, multimodal capabilities (e.g., voice or data visualization), and ensured domain knowledge on a larger span of financial topics. With the increasing adoption of AI-enabled advisory tools, businesses will be able to integrate this technology more effectively, to deepen customer engagement, increase operational efficiency, and drive financial inclusion, thus altering the fundamental way financial advice is imparted in the digital age.

Future work

In order to increase user trust and engagement, future developments in the Personalized Finance Chatbot will concentrate on enhancing explainability, risk management, and multi-modal capabilities. Explainable AI (XAI), one of the main areas for development, will enable the chatbot to offer clear explanations for suggested investments. Users will better grasp the rationale behind the recommendations of certain investment strategies by including risk assessment models, portfolio comparisons, and historical performance information. Furthermore, real-time market trend detection and suggestion refinement based on user input and economic conditions will be made possible by adaptive learning methods including reinforcement learning (RL) and sentiment analysis from financial news and social media.

Extending the chatbot's capabilities and adhering to financial standards is a crucial avenue for future research. By combining voice-based support and graphical data visualization with text-based interactions, the chatbot may help consumers engage with financial data in a more natural way. Maintaining adherence to international laws like the GDPR, SEC, and FINRA norms will be essential as AI-driven financial advising systems develop further. Trust in AI-powered financial services will be further increased by fortifying data security through authentication and encryption procedures. With the use of domain-specific LLMs like Fin BERT or Bloomberg GPT and hybrid AI techniques, further advancements in AI model development will allow the chatbot to comprehend financial jargon more efficiently and offer more profound market insights.

REFERENCES

1. Rivas, P., Zhao, L. [HTML][HTML] Marketing with chatgpt: Navigating the ethical terrain of gpt-based chatbot technology. (n.d.) Retrieved August 5, 2023, from www.mdpi.com/2673 2688/4/2/19
2. Rivas, P., Zhao, L. [HTML][HTML] Marketing with chatgpt: Navigating the ethical terrain of gpt-based chatbot technology. (n.d.) Retrieved August 5, 2023, from www.mdpi.com/2673 2688/4/2/19.
3. Kraiwanit, T., Jangjarat, K., & Atcharanuwat, J. (2022). The acceptance of financial robo-advisors among investors: The emerging market study. Journal of Governance and Regulation, 11(2, special issue), 332–339.10.22495/jgrv11i2siart12.
4. Abdulquadri A, Mogaji E, Kieu TA, et al. (2021) Digital transformation in financial services provision: a Nigerian perspective to the adoption of chatbot.
5. Sowa K, Przegalinska A, Ciechanowski L (2021) Cobots in knowledge work: Human–AI collaboration in managerial professions. J Bus Res 125:135–142.
6. Lai, K., & Ye, X. (2021). Privacy and Security Challenges in AI-Powered Financial Services.
7. Goel, A., & Sharma, A. (2021). Conversational AI in Banking and Finance: A Comprehensive Review. Journal of Banking and Financial Technology, 5(3), 243-260.
8. A. Moghar, M. Hamiche," Stock market prediction using LSTM recurrent neural network", Procedia Comp. Sc., vol. 170, pp. 1168 1173, 2020.
9. Toader DC, Boca G, Toader R, et al. (2020) The effect of social presence and chatbot errors on trust. Sustainability.
10. Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. NeurIPS 2020.
11. Buchanan, B. (2020). The Ethical Implications of AI in Financial Decision-Making. Harvard Kennedy School AI Policy Review.
12. McWaters, R., & Galaski, R. (2018). The new physics of financial services: Understanding how artificial intelligence is transforming the financial ecosystem. World Economic Forum.
13. Rajpurkar, P., Jia, R., & Liang, P. (2018). Know What You Don't Know: Unanswerable Questions for SQuAD. Proceedings of ACL 2018.
14. Aznoli F, Navimipour NJ (2017) Cloud services recommendation: Reviewing the recent advances and suggesting the future research directions. J Netw Comput Appl 77: 73–86. https://doi.org/10.1016/j.jnca.2016.10.009
15. Hegazy O, Soliman OS, Salam MA (2013 stock market prediction. Int J Comput Sci Telecommun (IJCST) 4(12)
16. Zhang,G.Peter.(2003)"TimeseriesforecastingusingahybridARIMAandneuralnetworkmode."Neurocomputing50:159-175.