JSS ACADEMY OF TECHNICAL EDUCATION BENGALURU

Department of Computer Science and Engineering

2021-2022

Naive Bayes Classification based Optimization

Name	USN	E-mail	Contact No.
Aditya Rastogi	1JS18CS008	adityarastogi1212@gmail.com	7667070675
Amit	1JS18CS015	1js18cs015@jssateb.ac.in	7897698725
Bharadwaj A	1JS18CS031	bharadwaj1878@gmail.com	8951802064
Jyothi L	1JS19CS403	jyothilokesh2001@gmail.com	9108847308

Under the Guidance of:

Mr. Sreenath M

Assistant Professor, Dept of CSE, JSSATE, Bengaluru

Introduction

Background

Question and Answer (Q&A) systems play an important role in our way of life for information and data sharing. Users post queries and choose inquiries to answer within the system. Thanks to the speedily growing user population and therefore the range of queries, it's unlikely for a user to come across a matter by accident that he/she will answer. Also, selflessness doesn't encourage all users to produce answers, to not mention top quality answers with a brief answer wait time. The project has been designed and enforced, a web social network primarily based Q&A system. It uses to its advantages the social network properties of common-interest associate degreed mutual-trust friend relationship to spot an utterer through relationship who will possibly answer the question, and enhance the user security.

Problem Statement

The Internet is a significant wellspring of data, where the measure of information is tremendous and continually developing. Clients depend on web search tools to discover explicit data in this information base. Web crawlers, for example, Google and Bing use catchphrases given by the clients to perform looks. As of late, mechanical innovative work exercises, for example, Microsoft and Facebook's social-highlighted Bing look attempt, endeavor to join web crawlers and online informal organizations for higher pursuit execution. As past research has shown, web indexes perform well in ordering website pages and furnishing clients with applicable substance to their look however are not appropriate for non-genuine inquiries, for example, "Which is the best nearby auto shop?". To address this specific class of non-authentic inquiries, many Question and Answer (Q&A) frameworks, for example, Yahoo! Answers, Baidu Zhidao, StackExchange, Quora and Ask have been created. Since their beginning, Q&A frameworks have demonstrated to be an important asset for sharing skill and thus are utilized by an enormous number of Internet clients. For instance, Yahoo! Answers was propelled toward the year's end 2005 and pulled in excess of 10 million clients in February , and hit 200 million

clients in December of 2009. Questions and answers frameworks additionally save all inquiries and answers, in this manner going about as a vault for data recovery. They are significant for sharing specialized information, yet in addition as a hotspot for accepting exhortation and fulfilling one's interest about a wide assortment of subjects. Accordingly, current Q&A frameworks may not meet the prerequisite of furnishing top notch answer with a short answer hold up time, however clients wish to get acceptable answers rapidly. This is affirmed by the examination in. It found that for Yahoo! Answers, just 17.6% of inquiries were addressed attractively; for the staying 82.4%, one fifth of the inquiries stayed unanswered. For Baidu Zhidao, 22.7% of inquiries were effectively replied, and 42.8% of the uncertain inquiries were not replied by any means.

Existing System

The developing significance of Q&A frameworks requests a push to all the more likely comprehend these frameworks and to improve. The works in concentrated the impact of various variables (e.g., clients' profiles, messages forecast, framework collaborations and network estimate) in the informal organizations on Q&A execution. These examination results establish the framework of this project to use interpersonal organization properties in the plan. Note that the current informal community dependent on the asker-answerer relationship in current Q&A frameworks is unique in relation to online interpersonal organization dependent on the social relationship, which is utilized in Naïve Bayes classifier optimization. The works in focused on finding specialists and definitive clients. Rather, Naïve bayes classifier optimization means to discover ordinary clients that can respond to questions including sentiment type questions. A few examinations have been led to make notoriety models in Q&A frameworks to build the believability of answers, and to decide the connection between the notoriety of the clients and the nature of their gave answers .Q&A system legitimately uses the interpersonal organization property of common trust companionship to persuade clients to give answers without depending on an extra notoriety model. Q&A system imparts closeness to other companion aide frameworks, for example, in utilizing the aggregate intensity of friends for a specific objective. Some exploration orders inquiries into predefined classifications,

making it simpler for clients to find recently posed inquiries and for specialists to discover addresses they can reply.

Proposed System

We have created and prototyped an online informal community based Q&A framework. It uses the properties of an interpersonal organization to advance an inquiry to potential answer suppliers, guaranteeing that a given inquiry gets a great answer in a brief timeframe. It expels the weight from answer suppliers by straightforwardly conveying them the inquiries they may be keen on, rather than requiring answer suppliers to look through an enormous accumulation of inquiries as in Yahoo! Answers or flooding an inquiry to the majority of an asker's companions in an online interpersonal organization. The sprout channel based improvement techniques scramble the intrigue and companionship data traded between clients to secure client protection, and record all n-grams of addressed inquiries to naturally recover responds to for intermittent inquiry. The onion steering based answer sending ensures the personalities of askers and answers.

Objectives

The main objective of Naïve Bayes classifier optimization is to boost the performance of Q&A systems by actively forwarding inquiries to users who are capable and want to answer the queries. It also aims to reduce wait time for the user to get answer to their queries.

System Requirements Specification

System requirements specification (SRS) is a nitty gritty detail of both the product and equipment segments required, which are important for the framework execution, alongside practical and non-useful and operational necessities, as foreseen from the framework.

Hardware requirements

Processor	Intel Core i5 or AMD FX 8 core series with clock speed of 2.4 GHz or above
RAM	2 GB or above
Hard disk	40 GB or above
Input device	Keyboard or mouse or compatible pointing devices
Display	XGA (1024*768 pixels) or higher resolution monitor with 32 bit color settings
Miscellaneous	USB Interface, Power adapter, etc.

Software requirements

Operating System	Windows XP or above
Programming Language	Python (3.6)
Integrated development environment (IDE)	Visual Studio Code
Algorithm Used	Naïve Bayes Classifier

Feasibility Study

Feasibility study is a significant stage in programming advancement process. It empowers the designers to have an appraisal of the item being created, as far as results of the item, its operational use and specialized help required for actualizing it. An investigation and assessment of a proposed venture to decide whether it is in fact plausible, is possible inside the evaluated expense, and will be gainful. Practicality ponders are quite often led where huge entireties are in question.

- Technical Feasibility
- Operational Feasibility
- Economic Feasibility

Technical Feasibility

The advancements required for your application is accessible in the market. Additionally the product can be actualized with existing technical assets. Technical feasibility is one of the principal ponders that must be directed after the venture has been recognized.

The Technical Feasibility Study computes materials required and evaluates the subtleties of how you will convey an item or administration. The Proposed System can be transferred to Cloud where every one of the calculations will occur. So it is a most ideal approach to contact all individuals since advanced mobile phones are so natural to deal with and even workstations are practically accessible to many.

Operational Feasibility

Operational feasibility is a proportion of how well a proposed framework takes care of the issues, and exploits the open doors recognized amid degree definition. The proposed application is easy to use with great GUI. The client just advances question to his companions, who has those interests. Henceforth, the created framework has operational plausibility.

Economical Feasibility

Economical Feasibility is utilized to decide the advantages and reserve funds that are normal from a proposed framework. Use acquired for building up the new framework will be practical or not. It is utilized to decide the advantages and reserve funds that are normal from proposed framework as this framework is open source it won't cost much. It requires just windows stage, for example workstations and PCs which are broadly utilized now days.

Algorithm used in Project

In AI we are regularly keen on choosing the best hypothesis (h) given information (d).

In a characterization issue, our hypothesis (h) might be the class to relegate for another information example (d).

One of the least demanding methods for choosing the most likely hypothesis given the information that we have that we can use as our prior learning about the issue. Bayes' Theorem gives a way that we can figure the probability of a hypothesis given our prior information.

Bayes' Theorem is expressed as:

$$P(h|d) = (P(d|h) * P(h))/P(d)$$

Where

- P(h|d) is the probability of hypothesis h given the information d. This is known as the back probability.
- P(d|h) is the probability of information d given that the hypothesis h was valid.
- P(h) is the probability of hypothesis h being valid (paying little mind to the information). This is known as the prior probability of h.
- P(d) is the probability of the information (paying little mind to the hypothesis).

You can see that we are keen on ascertaining the back probability of P(h|d) from the prior probability p(h) with P(D) and P(d|h).

In the wake of ascertaining the back probability for various theories, you can choose the hypothesis with the most elevated probability. This is the most extreme plausible hypothesis and may formally be known as the greatest a posteriori (MAP) hypothesis.

This can be composed as:

$$MAP(h) = max(P(h|d))$$

or on the other hand

$$MAP(h) = max((P(d|h) * P(h))/P(d))$$

or on the other hand

$$MAP(h) = max(P(d|h) * P(h))$$

The P(d) is a normalizing term which enables us to figure the probability. We can drop it when we are keen on the most plausible hypothesis as it is consistent and just used to standardize.

Back to characterization, in the event that we have a much number of cases in each class in our preparation information, at that point the probability of each class (for example P(h)) will be equivalent. Once more, this would be a steady term in our condition and we could drop it so we end up with:

$$MAP(h) = max(P(d|h))$$

Naive Bayes is an order calculation for double (two-class) and multi-class characterization issues. The system is most straightforward to comprehend when depicted utilizing paired or all out info esteems.

It is called naive Bayes or numbskull Bayes in light of the fact that the estimation of the probabilities for every hypothesis are streamlined to make their figuring tractable. As opposed to endeavoring to ascertain the estimations of each characteristic esteem P(d1, d2, d3|h), they are thought to be restrictively autonomous given the objective esteem and determined as P(d1|h) * P(d2|H), etc.

Packages Used

Scikit-Learn

Scikit-learn was at first created by David Cournapeau as a Google summer of code venture in 2007.

Scikit-learn creates a scope of administered and unsupervised learning calculations by means of a steady interface in Python. It is authorized under a lenient improved BSD permit and is dispersed under numerous Linux disseminations, empowering scholarly and business use. The library is based upon the SciPy (Scientific Python) that must be introduced before you can utilize scikit-learn. This stack incorporates:

- SciPy: Fundamental library for scientific computing
- IPython: Enhanced interactive console
- Matplotlib: Comprehensive 2D/3D plotting
- Sympy: Symbolic mathematics
- Pandas: Data structures and analysis
- NumPy: Base n-dimensional array package

Flask

Flask is a piece of the classifications of the smaller scale structure. Miniaturized scale system are typically structure with practically no conditions to outside libraries. This has advantages and disadvantages. Professionals would be that the system is light, there are little reliance to refresh and look for security bugs, cons is that some time you should accomplish more work without anyone else or increment yourself the rundown of conditions by including modules. On account of Flask, its conditions are:

- Werkzeug a WSGI utility library
- jinja2 which is its format motor

References

[1] Haiying Shen , Guoxin Liu , Haoyu Wang , Nikhil Vithlani. SocialQ&A: An Online Social Network Based Question and Answer System, 2017

[2] G. Drosatos, P. Efraimidis, A. Arampatzis, G. Stamatelatos, and I. Athanasiadis.

Pythia: A Privacy-enhanced Personalized Contextual Suggestion System for Tourism, 2015 .

[3] Sabrina Gaito, Elena Pagani ,Gian Paolo Rossi. Contact Analysis in Workplace,2009 .

[4] Matthew Richardson and Ryen W. White. Effects of expertise differences in synchronous social Q&A ,2012 .