Scientific Journal of Impact Factor (SJIF): 4.14

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 3, Issue 3, March -2016

Hybrid App Suite for Medical Diagnosis using Naive Bayes' Classifier

Praveen Hore, Assistant Professor,Army Institute of Technology, Pune
Anuj Tomar,Ashutosh Thakur,Nikhil Judge,Vishal Yadav
Department of Computer Engineering,Army Institute of Technology,Pune

Abstract—Medical and Healthcare industry comprises of sec- tor of industries that focus on patient well-being and medical research. These industries make up about 10% of the GDP of any developing nation. In India, the worth of medical industry is about $78 billion and it is expected to rise to $200 billion by 2020 [15] . IT industry has touched almost every domain of industries, medical industry remains to be exploited. IT industry will help in automation, better record management, and research in the healthcare industry. With the traditional methods, chances of medical negligence and misdiagnosis are prominent. We can use information technology to facilitate medical practitioners by providing them an interface to make calculations on-the-go, retrieve medical records and provide diagnosis by eliminating guesswork. By making calculations automated, degree of error is drastically reduced, which increases precision. Chances of medical record loss can be handled by maintaining a central server. Also, confidentiality of medical records can be achieved. In the current scenario, applications have to be developed for every platform separately. This consumes both time and money. Our solution proposes a hybrid application that will run on every platform. The application development is faster, simpler, and more rapid and the application is easier to maintain. You can change platforms anytime you need, Cordova lets you build your application for more than one platform just by one adding line of code.

Keywords—Hybrid application,Prediction,Cross-platform porta- bility

I. INTRODUCTION

Medical practitioners require a lot of complex calculations in which the amount of precision expected is very high. Such calculations might involve medical inputs such as sugar level, haemoglobin, etc. The first part of our app suite facilitates that, providing on-the-go calculations to medical practitioners. This can help them to make medical decisions.
The second part is related to medical diagnosis using Naive Bayes'classifier. Prediction can be done using this machine learning technique.
Both these parts come under the umbrella of a hybrid app suite which facilitates the use of the app on multiple platforms without the need of writing large pieces of extra code.

A. Motivation

The future for medical apps is bright. There will coexist both native and web based apps. App stores will continue to thrive and the market for consumer directed health apps will thrive. However, apps directed at physicians will become increasingly better connected to patients clinical records and will more likely be distributed by hospitals and other provider institutions.
According to industry surveys, by 2018, more than 1.7 billion smartphone and tablet users will have downloaded at least one mobile medical app (MMA) [16] . This level of acceptance and adoption of medical apps means new trends in innovation, and low cost medical services.These apps often control human physiology and work on sensitive health data, thus it is necessary to have evidences of their trustworthiness before actual marketing.

B. Literature Survey

Over the last fifteen years, IEC 62304 has become the benchmark standard for the development of medical device software, whether standalone software or otherwise, in both the EU and the US [17] . Leading industry innovation in software technologies has led key industry leaders and government regulators to recognise the emergence of numerous standalone medical software products that operates as medical devices. This has been reflected in regulatory changes. Example: Euro- pean medical devices directive MDD/93/42, updated in 2007. In Europe, a guidance document has been published on this subject. [18] The quality management system requirements for manufacturing a software medical

device, as is the case with any medical device, are described in the QSR regulations of the FDA and also in ISO 13485:2003. Software technology manufacturers that operate within the software medical de- vice space mandatorily develop their products in accordance with formal quality system requirements (cf. aforementioned FDA or EU requirements for this). Furthermore, though not mandatory, they frequently elect to obtain certification from a notified body, having implemented such quality system requirements as described within international standards, such as ISO 13485:2003.

A recent report Aitken co-authored , based on a study of nearly 43,700 purported health or medical apps available on Apple's iTunes app store, found that only 54% of them were "genuine" healthcare apps. Of those, 69% targeted consumers and patients while 31% were built for use by clinicians. Most of the consumer healthcare apps were "simple in design and do little more than provide information," according to the IMS report. Only 159 of the consumer apps could track or capture user-entered data, and fewer than 50 relate to condition management or provide tools and calculators for users to measure their vitals. That leaves "considerable room for growth in this sector," the report said [13] .

Aitken predicted that it will become commonplace for

patients to leave the doctorś office with a prescription for both a drug and an app and hospital discharge orders for patients will include downloading a mobile app on a patientś smartphone to provide an ongoing link between the provider and the patient. That will lead to better care and lower cost, he said [13] .

II. EXTERNAL INTERFACE REQUIREMENTS

A. *User Interface*

User have to log in with valid username and password in the hybrid app . If the details are correct then he will get access and will get logged in.Once he gets logged in all the details about patients will be displayed.

B. *Hardware Interface*

This involves interaction with the hardware by thye developed application. There are many reasons application might need to access hardware: Receiving mouse and keyboard events, accessing devices, such as a FireWire DV camcorder, and driving a device from an application are just a few. Although only code that resides in the kernel can access hardware directly.

C. *Communication Interface*

Communication interface is simple process to process interface. Mobile phone OS provide many services that allow us to communicate with hardware from plug-ins, applications, shared libraries, and other code running outside the kernel.

III. NON FUNCTIONAL REQUIREMENTS

A. *Reliability*

The quality of being reliable, dependable, or trustworthy. The quality of a measurement indicating the degree to which the measure is consistent, that is, repeated measurements would give the same result.

- Fault Tolerance
- Production Ready
- Graceful Failure
- Reliable Data Transfers

B. *Security*

It covers following issues :

- Secure access : Provides secure access to machine that protect both user and resources.
- Strong Authentication : Provides at least one strong authentication mechanism. Two step authentication is used.

C. Performance Requirements

The performance requirements may vary and the system will be subjected to a thorough analysis that gives a clear picture of its limitations.

D. Safety Requirements

All system data must be backed up every 24 hours and the backup copies stored in another server at different building or location for disaster recovery.

IV. ARCHITECTURE

The architecture of a Cordova hybrid-app contains follow- ing components:

- Web-app
- HTML Rendering Engine
- Cordova Plug-ins

A. Web-app

In computing, a web application or web app is a client- server application in which the client (or user interface) runs in a web browser. Web applications use web documents written in a standard format such as HTML and JavaScript, which are supported by a variety of web browsers. Web applications can be considered as a specific variant of client-server software where the client software is downloaded to the client machine when visiting the relevant web page, using standard procedures such as HTTP. Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, and acts as the universal client for any web application.

1) Interface: Through Java, JavaScript, DHTML, Flash, Silverlight and other technologies, application-specific meth- ods such as drawing on the screen, playing audio, and access to the keyboard and mouse are all possible. Many services have worked to combine all of these into a more familiar interface that adopts the appearance of an operating system.

2) Structure: Applications are usually broken into logical chunks called “tiers”, where every tier is assigned a role. Traditional applications consist only of 1 tier, which resides on the client machine, but web applications lend themselves to an n-tiered approach by nature.

B. HTML Rendering Engine

A web rendering engine is a software component that puts the marked up content(HTML, XML, image files, etc.) into formatted content, using formatting information, on the screen. It is divided into two parts for more modular approach:

- The engine does most of the work. It essentially takes a URL and a set of window content-area rect- angle coordinates as arguments. It then retrieves the document corresponding to the URL and paints a graphical representation of it in the given rectangle. It also handles links, forms,cookies, client-side script- ing,plugin,loading, and other matters.
- The host application provides the menu bar, address bar, status bar, bookmark manager, history and pref- erences functionality (among other things). It embeds the engine and serves as an interface between the user,

the engine, and the underlying operating system. Since it provides the graphical elements surrounding the area in which the engine paints documents, programmers sometimes use the term chrome to refer to its user interface (like the chrome surrounding a car) [19] .

V. STATISTICAL CLASSIFICATION USING MACHINE LEARNING

Supervised learning consists of a classifier and some training data. Based upon the training data and classifier the new/unseen data can be categorized.

A. Naive Bayes Classifier

In machine learning, nave bayes classifier are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong (naive) independence assumptions between the features.
Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable [12] . For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter . A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness and diameter features.

B. Probabilistic Model

Abstractly, naive Bayes is a conditional probability model [1]: given a problem instance to be classified, represented by a vector $p(C_k|x_1, ..., x_n)$ representing some n features (independent variables), it assigns to this instance probabilities for each of K possible outcomes or classes.
The problem with the above formulation is that if the number of features n is large or if a feature can take on a large number of values, then basing such a model on probability tables is infeasible [12] . We therefore reformulate the model to make it more tractable. Using Bayes' theorem, the conditional probability can be decomposed as

$$p(C_k|x) = \frac{p(C_k)p(x|C_k)}{p(x)}$$

In plain English, using Bayesian probability terminology, the above equation can be written as

$$posterior = \frac{prior \times likelihood}{evidence}$$

In practice, there is interest only in the numerator of that fraction, because the denominator does not depend on C and

the values of the features F_i are given, so that the denominator is effectively constant. The numerator is equivalent to the joint probability model

$$p(C_k|x_1, ..., x_n)$$

Which can be rewritten as follows, using the chain rule for repeated applications of the definition of conditional probability:

$$\begin{aligned} p(C_k, x_1, \ldots, x_n) &= p(C_k)\ p(x_1, \ldots, x_n|C_k) \\ &= p(C_k)\ p(x_1|C_k)\ p(x_2, \ldots, x_n|C_k, x_1) \\ &= p(C_k)\ p(x_1|C_k)\ p(x_2|C_k, x_1)\ p(x_3, \ldots, x_n|C_k, x_1, x_2) \\ &= p(C_k)\ p(x_1|C_k)\ p(x_2|C_k, x_1)\ \ \ldots p(x_n|C_k, x_1, x_2, x_3, \ldots, x_{n-1}) \end{aligned}$$

Now the "naive" conditional independence assumptions come into play: assume that each feature F_i is conditionally independent of every other feature for $F_j, j \neq i$, given the category . This means that

$$p(x_i|C_k,x_j) = p(x_i,C_k)\ \ p(x_i|C_k,x_j,x_q) = p(x_i,C_k)$$
$$p(x_i|C_k,x_j,x_q,x_l) = p(x_i,C_k)$$

And so on, for $i \neq j, q, l$, given the category C. Thus, the joint model can be expressed as

$$\begin{aligned} p(C_k|x_1, \ldots, x_n) &\propto p(C_k, x_1, \ldots, x_n) \\ &\propto p(C_k)\ p(x_1|C_k)\ p(x_2|C_k)\ p(x_3|C_k)\ \cdot\cdot \\ &\propto p(C_k) \prod_{i=1}^{n} p(x_i|C_k)\,. \end{aligned}$$

This means that under the above independence assumptions, the conditional distribution over the class variable is:

$$p(C_k|x_1, \ldots, x_n) = \frac{1}{Z} p(C_k) \prod_{i=1}^{n} p(x_i|C_k)$$

Where the evidence Z=p(x) is a scaling factor dependent only on $x_1, x_2, ..., x_n$, that is, a constant if the values of the feature variables are known.

C. Constructing a classifier from the probability model

The naive Bayes classifier combines this model with a decision rule. [14] One common rule is to pick the hypothesis that is most probable; this is known as the *maximum a*

posteriori or *MAP* decision rule. The corresponding classifier, a Bayes classifier, is the function that assigns a class label $\hat{y} = C_k$ for some k as follows:

C. Medscape

This app, by WebMD [10] , is another great medical reference tool offered on iOS and Android. The app is completely free, but does require registration for a free account (which you can do through the app itself) to use it. Once done, you can look up medications and drugs, check the disease reference tool, catch up on medical news, and much more.

1) Start.

VI. ALGORITHM

$$\hat{y} = \underset{k \in \{1,\ldots,K\}}{\operatorname{argmax}} \; p(C_k) \prod_{i=1}^{n} p(x_i | C_k).$$

VIII. CONCLUSION AND FUTURE WORK

This study has the objective to aid stakeholders in the current health care system.The combinations of attributes

2) Input doctor credentials.
3) If not correct re-direct else continue.
4) Input patient details.
5) Input patient symptoms.
6) Use classifier to calculate likelihood.
7) Display the likelihood of disease.
8) Display the precautions and medicines.
9) Exit.

VII. SIMILAR WORK

About 95 million Americans used their mobile phones either as healthcare tools or to find health information, according to a survey of more than 8,600 adults released in October by Manhattan Research, a New York-based healthcare marketing research firm. That's up 27% from 75 million in 2012. Smartphones have become an "indispensible" source of healthcare information for many, with 38% of smartphone users saying their device was "essential" for finding health and medical information. [13]

A. Epocrates

With more than 1 million active members (50 percent being U.S.-based physicians), Epocrates is being used by top doctors to look up drug information, find other providers for consults and referrals and quickly calculate patient measurements like BMI. Available for both iOS and Android, the app itself and most of its content is free, access to additional information and functionality (like lab guides, alternative medications, and disease information) requires an in-app purchase of Epocrates Essentials for $159.99 a year.

B. Doximity

Doximity is the largest medical professional network in the U.S., with over 40 percent of physicians as members. Through both mobile and web platforms, physicians can use Doximity's free tools for HIPAA-secure communication, electronic faxing, reading custom-curated medical news, and career management. Doximity is designed exclusively for

healthcare professionals. With both iOS and Android versions, the app is free to download, but does require you to sign up for membership in the network.
selected as the best predictors of the selected target attributes,
which themselves represent a measure of treatment outcome or a proxy for cost, might have utility for stakeholders in the current health care system. Several types of stakeholders make decisions based on this type of information. The best combinations of attributes might have planning implications for hospital administrators, treatment protocol implications for physician groups, and public health implications for legislators, government agencies, and think tanks. The combination of a classifier for prediction of ailment with on-the-go calculations did consistently identify the most-predictive attributes for a given target attribute, although greater than modest increases in classification accuracy would have been gratifying. For each choice of target attribute, the most-predictive attributes, alone or in combination, were those relating to diagnosis.
By making calculations automated, degree of error is drastically reduced, which increases precision. Chances of medical record loss can be handled by maintaining a central server. Also, confidentiality of medical records can be achieved.
Future work includes a more accurate method of predicting the ailment. Also, the medical calculations can be made more interactive by combining the application with the suitable hardware. This can eliminate the need for the manual input of data, and thus, can provide more accuracy.

REFERENCES

[1] Stephanie J. Hickey, Iona College, USA. Communications of the IIMA(2013). *Naive Bayes Classification of Public Health Data with Greedy Feature Selection*

[2] Abraham, R., Simha, J. B., & Iyengar, S. S. (2006). *A comparative analysis of discretization methods for medical data-mining with naive Bayesian classifier*. In S. P. Mohanty & A. Sahoo (Eds.) Proceedings of the 9th International Conference on Information Technology (pp. 235-236). Bhubaneswar, India: IEEE Computer Society. doi: 10.1109/ ICIT.2006.5

[3] Ahuja, H. ; Johari, R.,*Optimization of the issues in the migration from Android Native to Hybrid Application: Case study of Student's Portal application* Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference)

[4] Abraham, R., Simha, J. B., & Iyengar, S. S. (2007). *Medical datamining with a new algorithm for feature selection and naive Bayesian classifier*. In S. K. Rath & P. Mohapatra (Chairs) 10th International Conference on Information Technology (pp. 44-49). Rourkela, India: IEEE. doi: 10.1109/ICIT.2007.41

[5] Al-Aidaroos, K. M., Bakar, A. A., & Othman, Z. (2012). *Medical data classification with naive Bayes approach.* Information Technology Journal, 11, 1166-1174. doi: 10.3923/itj.2012. 1166.1174

[6] Bellazzi, R., & Zupan, B. (2008). *Predictive data mining in clinical medicine: Current issues and guidelines.* International Journal of Med- ical Informatics, 77(2), 81-97.

[7] Cooper, G. F., Aliferis, C. F., Ambrosino, R., Aronis, J., Buchanan,
B. G., Caruana,(1997). *An evaluation of machine-learning methods for predicting pneumonia mortality*. Artificial Intelligence in Medicine, 9(2), 107-138.

[8] Domingos, P. (2012). *A few useful things to know about ma- chine learning.* Communications of the ACM, 55(10), 78-87. doi: 10.1145/2347736.2347755

[9] Hassan, S. Z., & Verma, B. (2007). *A hybrid data mining approach for knowledge extraction and classification in medical databases*. In L. M. Mourelle, N. Nedjah, J. Kacprzyk, & A.Abraham (Eds.) Seventh In- ternational Conference on Intelligent Systems Design and Applications (pp. 503-510). Rio de Janerio, Brazil. IEEE Computer Society.

[10] *WebMD- medical information website*, http://www.webmd.com

[11] Huang, Y., McCullagh, P., Black, N., & Harper, R. (2007). *Feature selection and classification model construction on type 2 diabetic patients' data.* Artificial Intelligence in Medicine, 41(3), 251-262.

[12] http:// www.wikipedia.org/wiki/Naive Bayes classifier

[13] http://www.modernhealthcare.com/article/20131214/MAGAZINE/312149983 [14] Yuanshun Dai, Yanping Xiang, Gewei Zhang (2009). *Self-healing and*
Hybrid Diagnosis in Cloud Computing

[15] ASA Associates LLP, (2015) *A Brief Report on Health- care,Telemedicine Medicine in India*

[16] FDA. Mobile medical applications, http://www.fda.gov/medicaldevices/digitalhealth/mobilemedicalapplications/default.htm

[17] U.S. Department Of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health and Center for Biologics Evaluation and Research (11 January 2002). *General Principles of Software Validation; Final Guidance for Industry and FDA Staff*

[18] Medical Device Certification, Stuttgart, Germany, *Basic Information about the European Directive 93/42/EEC on Medical Devices*

[19] https://en.wikipedia.org/wiki/Web_browser_engine

