A Decision Support System for Pediatric Diagnosis

Precious Iheme, Nicholas Omoregbe, Sanjay Misra(✉), Foluso Ayeni, and Davies Adeloye

Covenant University, Ota, Nigeria
{Precious.Iheme,Nicholas.Omoregbe,Foluso.Ayeni,Sanjay.Misra, Davies.Adeloye}@covenantuniversity.edu.ng

Abstract. Newborns are fragile and have a high risk of dying within the first 28 days of their life, therefore they require quality care from conception. This research aims at implementing a mobile pediatric diagnostic system for the rural settlers in Nigeria, reducing childhood mortality and providing an alternative pediatric professional. 581 records classified with naïve Bayes and decision-stump-tree classifier gave a higher accuracy level for naïve Bayes. A decision-support system is developed to aid health workers in rural areas in providing quality health service for children below six, which will provide low-cost medical service and contribute to reducing childhood mortality.

Keywords: Child healthcare · Mobile technology · Naïve bayes Pediatric disease

1 Introduction

Newborns are highly prone to health risk due to their fragile nature; the risk of a child dying is high during the first 28 day of their life in several developing countries including Africa [1]. Hence, they require quality care from the time of conception to the post-natal stage and further. About 6.3 million children under the age of 5 died in 2013, preterm birth, delivery complications and infections source a great number of neonatal deaths [2–5]. These early child deaths are due to conditions that could be prevented or treated with access to simple, affordable interventions [6–8]. Children in sub-Saharan Africa are more than fourteen times more likely to die before they get the age of 5 than children in developed regions. In Nigeria about 2300 children under age five die each day, the rural areas recording a larger number of infant mortality. One of the SDGs goals (sustainable development goals) is to end preventable deaths of newborns and children less than 5 years of age [9–11]. UNICEF also, in partnership with the Nigerian government to improve the quality of healthcare services as well as decrease the rate of infant mortality [12].

Mobile technology has improved the quality of health care services, especially in developing countries [13, 14]. In many places around the world, there has been an explosion of mobile apps, remote monitoring devices, and online instructional materials [15, 16]. This has brought updated and constant information to health assistants such as midwives who are on the forefronts of care and assisting rural populations who need

C. M. F. Kebe et al. (Eds.): InterSol 2017/CNRIA 2017, LNICST 204, pp. 177–185, 2018.
https://doi.org/10.1007/978-3-319-72965-7_17

medical treatment and advice [17, 18]. Hence incorporating the goals of UNICEF and SDGs into a mobile application will make those services readily accessible to its target especially in the rural areas.

Development of such application requires data gathering, preprocessing, processing (data mining) and evaluation (as shown in Fig. 2) before it can be deployed as knowledge. Because of uncertainty and incomplete information faced when designing automated healthcare systems [19], naive Bayes classifier which has been tested and trusted to handle incomplete data successfully despite its simplicity [20–23]. In the present work we have developed a decision support system using WEKA tools. We have also developed a mobile application by using our own algorithms [24] where we also compared the results obtained by applying WEKA tool with our own system.

The rest of this paper is organized as follows; Sect. 2 give the statement of the problem and objective of the study, Sect. 3 shows a review of related findings and technologies, Sect. 4 shows the Research methodology, Sect. 5 shows the validation of the system, Sect. 5 shows the result and discussion, Sect. 6 shows the conclusion and future work.

2 Statement of Problem and Objective of Study

Low pediatrician to patient ratio which has contributed to the increase of infant mortality, hence alternative measures are required to improve healthcare service. Also the development of decision support systems has contributed to easy accessibility to quality, healthcare services in another country especially the developed country. The implementation of this system in Nigeria, especially for the rural settlers, will reduce childhood mortality and morbidity and well as bridge the gap of the distribution of medical professional within the six geopolitical zones, especially those with the low pediatrician to patient ratio. The research aims to implement a decision support pediatric diagnostic system using the naïve Bayes classifier particularly for under-fives, and a mobile expert system using this model is proposed towards improving the quality of health services available for this age group in Nigeria.

This work gathered requirements, standard symptoms, case definitions and diagnostic criteria of some common childhood diseases. It modeled core functionalities of the system. A naive Bayesian framework model for learning the standard symptoms, case definitions, and diagnostic criteria as well as inference making was designed. A prototype using the Bayesian framework was also designed. The main objective of this work is to explore the usefulness of data mining techniques and Decision Support technologies on Healthcare Systems.

3 Brief Review of Related Findings/Technologies

The advent of mobile technology has encouraged the design of complex systems in a simple mobile application. The advantage of these systems cannot be overemphasized, from the bridging distance between experts and clients to creating an avenue for interns to learn amongst many others.

Support vector machine was used in [25] to identify the particular heart valve disease that a patient is diagnosed with. It was also used by [26] to develop an automatic medical diagnostic system using statistical medical information provided by the user.

Decision tree classification algorithm was used to extract from laboratory finding the distinguishing factor between a cognitive heart failure and those of the dyspnea class [27]. In [28], it was used to prevent whether cancer is benign or malignant using the 10-fold cross validation test.

The decision tree C4.5, bagging with decision tree C4.5, and bagging with a naive Bayes classification algorithms was compared in the identification of heart diseases in patients using the 10 fold validation to compute the confusion matrix [29].

Support Vector Machines (SVM) and naïve Bayes classifier were compared for text categorization wikitology. Naïve Bayes prove to be better with 28.78% improvement using the 10-fold validation [30].

Nematzadeh compared decision tree with naïve Bayes method in the classification of researcher's cognitive styles in an academic environment. Naïve Bayes produced a better accuracy with 93.83637 [31].

Different artificial intelligence methods have been used to design decision support systems all with their individual advantages depending on the nature of data obtained. In [32] the self-organizing fuzzy neutral network with truth-value fuzzy inference was used in the classification of acute lymphoblastic leukemia (ALL) subtypes using gene expression data.

4 Research Methodology

In order the achieve the objectives of this research, existing patient record on some common childhood diseases were observed, preprocessed, evaluated and compared with another classifier. Diagnostic criteria were obtained through interaction with doctors, reviewing WHO data to identify standard symptoms, case definitions and diagnostic criteria of some common childhood diseases, about 1107 data on symptoms and diagnosis from patient notes of children between 0–5 were obtained.

The data mining architecture was used to model the system on a mobile platform. The system works offline and uses dialog technique via SMS to obtain answers from the user.

4.1 System Architecture

Data mining architecture is made up of the business and data layers including the user interface (as depicted in Fig. 1).

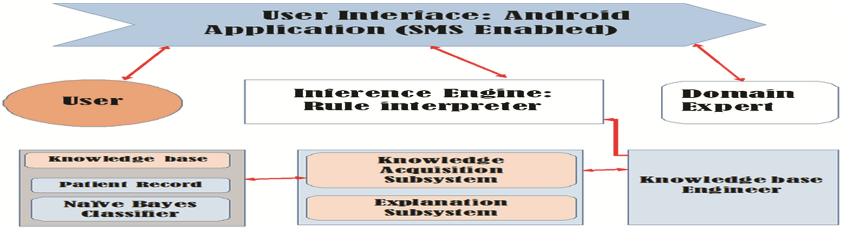

Fig. 1. System architecture

The user interface requires dialog between the user and the system. The choice of the user determines the tool that will be used. The business component involves data processing, which is achieved through four major tasks namely: preparation and preprocessing, processing (data mining), Evaluation and knowledge deployment. The data is stored in a data warehouse where persistent data can be extracted [33].

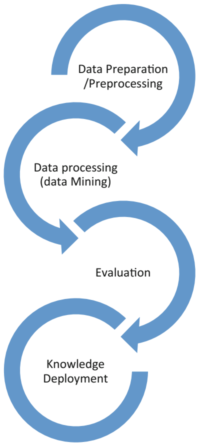

Fig. 2. Paediatric application developmental stages

The data-mining algorithm used is the Naïve Bayes classifier. The model learns a set of preprocessed data collected and the output is classified based on the different diseases of focus for inference purpose. The framework is then integrated into a mobile platform for easy accessibility of the system to target audience.

4.2 Naïve Bayes Classifier

A classifier is a function that maps input features to outclass label. Its goal is to learn from a labeled training set of input-output pairs for predictive purpose [34]. Naïve Bayes classifier had its name from its naïve assumption of independence between the different features of class label.

The Bayesian concept is explained thus;

A tuple X belong to a class C. X has a measurement made on a set of attribute n. Given a hypothesis H, classification problem determines that the hypothesis holds given evidence (H/X). P(H/X) is a posterior probability e.g. the probability that patient X with attributes fever and vomiting will belong to a class malaria. P(H) is a prior probability e.g. the probability that any patient will have malaria. P(X) is a prior probability e.g. the probability that a fever with fever has malaria.

$$\text{Bayes theorem: } (P(H)/(X)) = (P(X/H)P(H))/(P(X)) \tag{1}$$

The steps taken to achieve classification in naïve Bayes are stated below:

Step 1: D is a training set and its class variables. Each tuple X is represented by an n-dimensional attribute vector ($X = x_1, x_2 \ldots x_n$) with n measurement on the n attribute ($a_1, a_2 \ldots a_n$). Hence the data to be used for classification must be presented in the format understandable by classier (usually in a tabular, comma separated or Attribute-Relation file format (ARFF) format).
Step 2: The classifier calculates the prior probabilities for each class. That is, Given m classes C1, C2,…, Cn and a tuple X, the classifier will predict that X belong to the class with the highest posterior probability, conditioned on X.

$$P(C_i/X) > P(C_i)\,P(X) \text{ for } 1 \le j \le m,\ j \ne 1 \tag{2}$$

Step 3: since P(X) is constant for all the class, only P(Ci/X) P(Ci) needs to be maximized. $P(C_i) = C_{iD}/D$.
Step 4: To reduce computation in evaluating $P(X/C_i)$, the naïve Bayes assumption of class-conditional independence is made.

$$P(X/Ci) = \prod_{(k=1)}^{n} P(X_k/C_i)\ \left(P(X_1/C_i),\ P(X_2/C_i) \ldots\ P\ (X_n/C_i)\right) \tag{3}$$

Step 5: To predict the class label of X, $P(X/C_i)\ P(C_i)$ is evaluated for each class C_i.

$$P(X/C_i)\ P(C_i) > P(X/C_j)\ P(C_j) \text{ for } 1 \le j \le m,\ j \ne i \tag{4}$$

5 Validation and Discussion of the Model

After preprocessing have been carried out on the data collection 581 were selected after removing irrelevant and missing attributes. The data was evaluated using the WEKA tool for classification.

The result from the classification using the 10 fold cross validation test option, shows that 333 of the dataset were correctly classified whereas 247 were incorrectly classified gave 57.4% accuracy, with a means square error of 0.3929. Malaria had the higher incorrectly classified instances (92 instances) while measles has the least incorrectly classified instances (i.e. 14 instances) (Fig. 3).

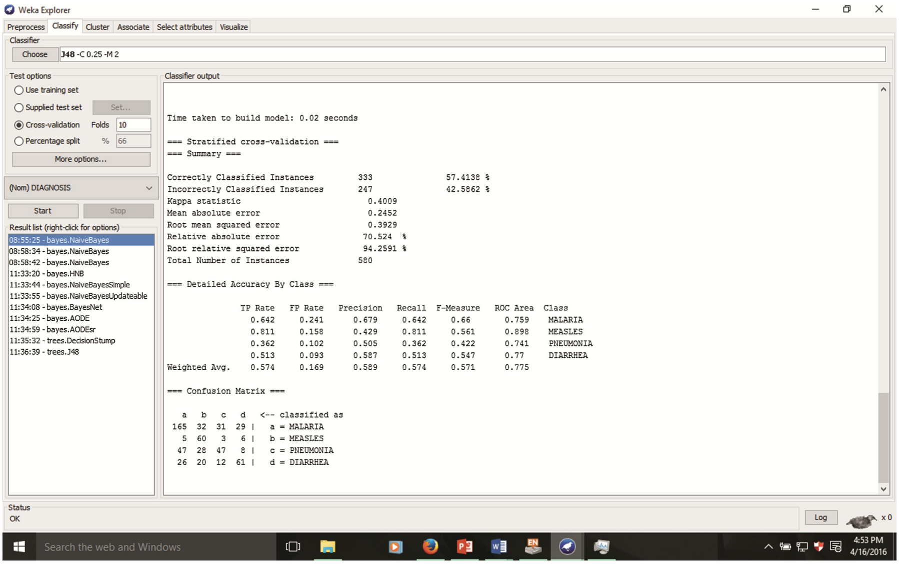

Fig. 3. Result of data evaluation using the 10 fold validation

Using the 66% split option the accuracy obtained was 55.3% (as shown in Fig. 4). 109 instances were classified correctly, 88 were incorrectly classified with a means square error of 0.3998. Malaria had the higher incorrectly classified instances (28 instances) while measles has the least incorrectly classified instances (6).

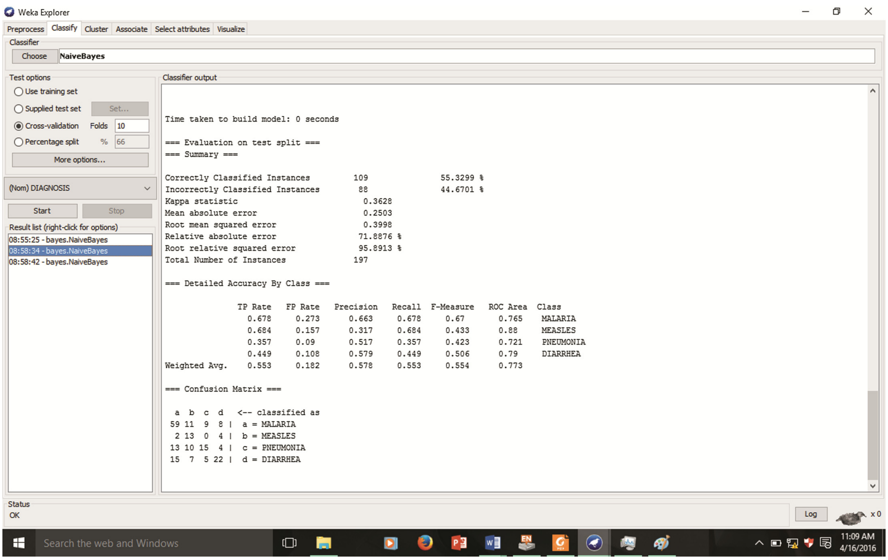

Fig. 4. Result of data evaluation using the 66% split

The result of the Naïve Bayes classifier was compared with that of the decision stump tree classifier, the result obtained was lower than the accuracy level from the naïve Bayes classifier.

The result from the decision stump tree (as shown on Fig. 5) reveals that 285 instances were classified correctly, 295 were incorrectly classified, with a means square error of 0.4026. Pneumonia had the higher incorrectly classified instances (130 instances) while malaria has the least incorrectly classified instances (55 instances).

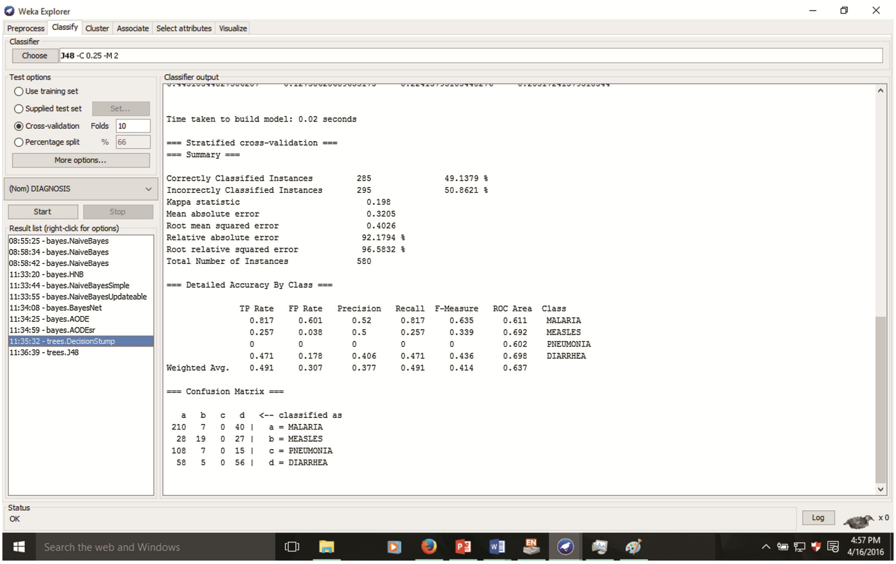

Fig. 5. Result for data evaluation using the decision stump tree

The result obtained will be suitable to design a decision support system. This system will require verification of diagnosis by an expert. However, it will provide an avenue for interns and medical technician to have more knowledge of the diseases diagnosed by the system.

6 Conclusions and Future Work

From this research, a decision system was developed to aid health workers in the rural areas in Nigeria in providing quality health service for children below age six. One advantage of this system is that it is SMS-based; hence the patients can receive medical advice from the comfort of their homes. This also reduces the number of the patient a doctor will have to see in a day and reduce long waits at the hospitals and delayed diagnosis, and consequently death.

This research is significant as it provides a low-cost medical service to children especially in the rural regions, contribute to reduction of childhood mortality through accessible medical service, educate medical personnel in the rural areas of diagnosis of childhood disease, and make data available to doctors for further research as well as bridging the communication gap between doctors and patients.

This research is a contribution to the achievement of SDG's goal to reduce childhood mortality between 2016 and 2030 and UNICEF's Accelerating Child Survival and Development in rural areas.

Acknowledgement. We acknowledge the support and sponsorship provided by Covenant University through the Centre for Research, Innovation and Discovery (CUCRID).

References

1. UN-DESA.: Transforming our world: the 2030 Agenda for Sustainable Development (2015). https://sustainabledevelopment.un.org/post2015/transformingourworld
2. Chan, M., Lake, A.: Towards ending preventable child deaths. Lancet 379 (2012)
3. Wardlaw, T., You, D., Newby, H., Anthony, D., Chopra, M.: Child survival: a message of hope but a call for renewed commitment in UNICEF report. Reprod. Health **10**, 1–4 (2013)
4. George, C.M., Vignola, E., Ricca, J., Davis, T., Perin, J., Tam, Y., Perry, H.: Evaluation of the effectiveness of care groups in expanding population coverage of Key child survival interventions and reducing under-5 mortality: a comparative analysis using the lives saved tool (LiST). BMC Publ. Health 1–9 (2015)
5. Shifa, G.T., Ahmed, A.A., Yalew, A.W.: Early days of life are crucial for child survival in gamo gofa zone, Southern Ethiopia: A community based study. BMC Pediatr. **16**, 1–10 (2016)
6. Hershey, C.L., Doocy, S., Anderson, J., Haskew, C., Spiegel, P., Moss, W.J.: Incidence and risk factors for malaria, pneumonia and diarrhea in children under 5 in UNHCR refugee camps: A retrospective study. Confl. Health **5**, 1–11 (2011)
7. Kahabuka, C., Kvåle, G., Hinderaker, S.G.: Factors associated with severe disease from malaria, pneumonia and diarrhea among children in rural Tanzania – A hospital-based cross-sectional study. BMC Infect. Dis. **12**, 1–9 (2012)
8. Yakoob, M.Y., Theodoratou, E., Jabeen, A., Imdad, A., Eisele, T.P., Ferguson, J., Jhass, A., Rudan, I., Campbell, H., Black, R.E., Bhutta, Z.A.: Preventive zinc supplementation in developing countries: impact on mortality and morbidity due to diarrhea, pneumonia and malaria. BMC Publ. Health **11**, 1–10 (2011)
9. WHO: Health in 2015: from MDGs to SDGs (2016). http://www.who.int/gho/publications/mdgs-sdgs/en/
10. Sachs, J.D.: Happiness and Sustainable Development: Concepts and Evidence. World Happiness
11. Ahmed, M.: The challenges ahead. Education (2015)
12. UNICEF: Accelerating Child Survival and development. The Nigerian Child: United Nations Children's Fund Nigeria Newsletter, pp. 1–12 (2008)
13. Sa, J.H.G., Rebelo, M.S., Brentani, A., Grisi, S.J., Iwaya, L.H., Simplicio Jr., M.A., Carvalho, T.C.M.B., Gutierrez, M.A.: Georeferenced and secure mobile health system for large scale data collection in primary care. Int. J. Med. Inform. **94**, 91–99 (2016)
14. Sun, N., Rau, P.L.P., Li, Y., Owen, T., Thimbleby, H.: Design and evaluation of a mobile phone-based health intervention for patients with hypertensive condition. Comput. Hum. Behav. **63**, 98–105 (2016)
15. West, D.M.: Using mobile technology to improve maternal health and fight Ebola: A case study of mobile innovation in Nigeria. Center for Technological Innovation at Brookings (2015)

16. Hampshire, K., Porter, G., Owusu, S.A., Mariwah, S., Abane, A., Robson, E., Munthalie, A., DeLannoy, A., Bangog, A., Gunguluzaf, N., Milnere, J.: Informal m-health: How are young people using mobile phones to bridge healthcare gaps in Sub-Saharan Africa? Soc. Sci. Med. **142**, 90–99 (2015)
17. Uddin, M.J., Shamsuzzaman, M., Horng, L., Labrique, A., Vasudevan, L., Zeller, K., Chowdhury, M., Larson, C.P., Bishai, D., Alamj, N.: Use of mobile phones for improving vaccination coverage among children living in rural hard-to-reach areas and urban streets of Bangladesh. Vaccine **34**(2), 276–283 (2016)
18. Domek, G.J., Contreras-Roldan, I.L., O'Leary, S.T., Bull, S., Furniss, A., Kempe, A., Asturias, E.J.: SMS text message reminders to improve infant vaccination coverage in Guatemala: A pilot randomized controlled trial. Vaccine **34**(21), 2437–2443 (2016)
19. Nada, N., Adepa, I., Fatin, G.: Design a fuzzy expert system for pediatrics diseases diagnosis. Raf. J. Comp. Maths **5**(2), 155–173 (2008)
20. Corani, G., Zaffalon, M.: JNCC2: An extension of naive Bayes classifier suited for small and incomplete data sets. Environ. Model. Softw. **23**(7), 960–961 (2008)
21. Soria, D., Garibaldi, J.M., Ambrogi, F., Biganzoli, E.M., Ellis, I.O.: A 'non-parametric' version of the naive bayes classifier. Knowl. Based Syst. **24**(6), 775–784 (2011)
22. Chen, J., Huang, H., Tian, F., Tian, S.: A selective bayes classifier for classifying incomplete data based on gain ratio. Knowl. Based Syst. **21**(7), 530–534 (2008)
23. Bounhas, M., Hamed, M.G., Prade, H., Serrurier, M., Mellouli, K.: Naive possibilistic classifiers for imprecise or uncertain numerical data. Fuzzy Sets Syst. **239**, 137–156 (2014)
24. Iheme, P., Omoregbe, N., Misra, S., Adeoye, D., Adewole, A.: Mobile-Bayesian diagnostic system for childhood infectious diseases. In: Proceedings of ICADIWT 2017, Maxico (2017)
25. Maglogiannis, I., Loukis, E., Zafiropoulos, E., Stasis, A.: Support vectors machine-based identification of heart valve diseases using heart sounds. Comput. Methods Programs Biomed. **95**(1), 47–61 (2009)
26. Kampouraki, A., Vassis, D., Belsis, P., Skourlas, C.: e-Doctor: A web based support vector machine for automatic medical diagnosis. Procedia Soc. Behav. Sci. **73**, 467–474 (2013)
27. Son, C.S., Kim, Y.N., Kim, H.S., Park, H.S., Kim, M.S.: Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches. J. Biomed. Inform. **45**(5), 999–1008 (2012)
28. Sumbaly, R., Vishnusri, N., Jeyalatha, S.: Diagnosis of breast cancer using decision tree data mining technique. Int. J. Comput. Appl. **98**, 1–9 (2014)
29. Tu, M.C., Shin, D., Shin, D. (eds.): A comparative study of medical data classification methods based on decision tree and bagging algorithms. In: Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing (2009)
30. Hassan, S., Rafi, M., Shaikh, M.S.: Comparing SVM and naive bayes classifiers for text categorization with Wikitology as knowledge enrichment. In: IEEE 14th International Multitopic Conference (INMIC) (2011)
31. Nematzadeh, B. Z.: Comparison of Decision Tree and Naive Bayes Methods in Classification of Researcher's Cognitive Styles in Academic Environment (2012)
32. Tung, W.L., Quek, C.: GenSo-FDSS: a neural-fuzzy decision support system for pediatric ALL cancer subtype identification using gene expression data. Artif. Intell. Med. **33**(1), 61–88 (2005)
33. Winsemann, T., Köppen, V.: Persistence in enterprise data warehouses. Otto-von-Guericke University Magdeburg, Technical Reports (2012)
34. Murphy, K.P.: Naive bayes classifiers. University of British Columbia (2006)