临床医学研究: 2022 年 4 卷 12 期
ISSN: 2705–0939(Print); 2705–0475 (Online)

# 埃塞俄比亚育龄妇女破伤风类毒素免疫获取的数据挖掘

凯迪尔・侯赛因・阿贝加兹[1*]　埃米鲁・梅尔达萨・阿托姆萨[2]
1.马达瓦拉布大学卫生科学学院公共卫生系生物统计和卫生信息学 埃塞俄比亚 贝尔戈巴 999047
2.生物统计和卫生信息学 西沃勒加地区卫生部 埃塞俄比亚 奥罗米亚金比 999047

【摘　要】：破伤风类毒素（TT）疫苗是为预防新生儿破伤风和破伤风引起的孕产妇死亡而接种的育龄妇女疫苗。在全球范围内，破伤风每年造成 5%的产妇死亡和 14%的新生儿死亡。数据挖掘是从大量数据中发现有趣的模式和知识的过程。因此，本研究的目的是识别最佳分类器，并使用数据挖掘算法技术从 TT 数据集中预测模式。本研究的数据来自 2011 年埃塞俄比亚人口与健康调查（EDHS）的破伤风类毒素数据集，并使用选择、处理、转换、挖掘和解释的知识发现过程进行分析。使用 WEKA 3.6.1 工具进行分类、聚类、关联和属性选择。分类器在训练数据上的准确率相对高于测试数据，多层感知器是我们的破伤风类毒素数据集中最好的分类器。在 10 倍交叉验证中，正确分类最好的是 naïve 贝叶斯法 63.30%，最近邻法 60.52%准确率最低。使用 Naïve Bayesian 进行单个数据实例测试，创建测试 1、测试 2、测试 3 和测试 4 数据测试实例，其中三个数据被正确预测，但其中一个被错误分类。在一般联想中获得的最大置信度为 0.98。但是，在 class 属性中，它是 0.72。母亲识字率的信息增益较高，为 0.046。综上所示，基于 TT 疫苗接种数据的最佳算法为多层感知器分类器，准确率为 67.28%，构建模型总时间为 0.01 秒。多层感知器分类器的平均误差最小，为 32.72%。这些结果表明，在测试的机器学习算法中，多层感知器分类器有潜力显著改进用于破伤风类毒素 EDHS 数据的传统分类方法。
【关键词】：数据挖掘；WEKA；分类；聚类；破伤风类毒素（TT）；EDHS

## Data Mining of Access to Tetanus Toxoid Immunization Among Women of Childbearing Age in Ethiopia

**Kedir Hussein Abegaz[1*]　Emiru Merdassa Atomssa[2]**

1.Biostatistics and Health Informatics Public Health Department College of Health Sciences Madda Walabu University Ethiopia Bale Goba 999047

2.Biostatistics and Health Informatics West Wollega Zonal Health Department Ethiopia Gimbi Oromia 999047

**Abstract:** Tetanus toxoid (TT) vaccine is given to women of childbearing age to prevent neonatal tetanus and maternal mortality attributed to tetanus. Globally, tetanus is responsible for 5% of maternal deaths and 14% of neonatal deaths annually. Data mining is the process of discovering interesting patterns and knowledge from large amounts of data. Thus, the aim of this study was to identify the best classifier, and to predict the pattern from the TT data set using the data mining algorithms technique. The data for this study were the Tetanus Toxoid data set from the Ethiopian Demographic and Health Survey (EDHS) 2011, and analyzed using the Knowledge discovery process of Selection, Processing, Transforming, mining, and interpretation. The WEKA 3.6.1 tool was used for classification, clustering, association and attribute selection. The accuracy rate of the classifiers on training data is relatively higher than on test data and the multilayer perceptron is the best classifier in our data set on Tetanus toxoid. In the cross-validation with 10 folds, correctly classified best are by naïve Bayesian 63.30% and the least accurate were by k-nearest neighbor 60.52%. Single data instance test using Naïve Bayesian was done by creating test 1, test 2, test 3, and test 4 data test instance, three of them are correctly predicted but one of them incorrectly classified. The maximum confidence attained in the general association is 0.98. But, in the class attribute, it is 0.72. The literacy status of the mother has high information gain with the value 0.046. As a conclusion, the best algorithm based on the TT vaccination data is multilayer perceptron classifier with an accuracy of 67.28% and the total time taken to build the model is at 0.01 seconds. Multilayer perceptron classifier has the lowest average error at 32.72% compared to others. These results suggest that among the machine learning algorithm tested, multilayer perceptron classifier has the potential to significantly improve the conventional classification methods for use in EDHS data of Tetanus toxoid.
**Keywords:** Data Mining; WEKA; Classification; Clustering; Tetanus Toxoid (TT); EDHS

## 1 简介

┳ 岡斃俓佶卻婇娐♗揩孌嗸學曁TT暩鮀巃朁▾ 拖捨吣擭

キ婷♗揩丆伶婷♗揩傷崠她、俓鳴╩ ⇑ 伉忤俩㎘僈┫ 份傻朁
倞キ吹咟鳴╩ 她┳ 徥㉱伶凋凋呖ㄛ倏呀噔咕憹健㈦擭嫯卩⇑

伶嗔巹偓偌卻㌘婷ㅿ撍孌嶅學ㅃ扉巹▾ こ勿俌◆ 她丅壚シ㌘吣壚キ婷ㅿ撍她凅中[1]⇑ 婷ㅿ撍呖壯婷ㅿ撍商帇㊀喽壚抄慮婁┦ 、壚她嶅學冢慞她⇑ 劉哯呖慄慮伉墆佚┦ 卻徹◑ ⊡婷卞她妄岘鞅嗚╩ 宦宪暨俏ㅿ㌢瞭鞅冶峻克影さㄷ寀帇佩偰呀昆澤她⇑ ┘ 壼㈦壚宦宪♠忓巹㈲【咾崠妒嚭葊寁㌾悽她份傻巹吣壚キ婷ㅿ撍暨NNT瞭疥〕勻kΩ塺╋▌┳ 5%[2]⇑

伉ダ増蕘伻ピ巹嗷葅咕 5%她偈、偓嗚╩ 丆 14%她吣壚キ嗚╩ 呖壯婷ㅿ撍冢慞她巹伉┆╥ 揓嚞份傻巹吣壚キ嗚╩ 塺揮悽 25%[3-5]⇑ 勧崟 2012 葅 12 咔巹伉 30┤ 份傻巹┳ 徥呖伉揓嚞丆╥ 嚞巹、偓丆吣壚キ婷ㅿ撍┆ 旻呖┆ ┤ チテ㈦壚揤揺 [2]⇑ 伉収个匔▾ ㈳揓嚞巹♠忓嗷葅咕俩悽 70000 fm吣壚キ伉丅壚nm她侯伋log嗚╝ 吣壚キ婷ㅿ撍[5]⇑ 住佺え嗺╥ 呖┘ 壼┐吣壚キ婷ㅿ撍㌾疥塺丆嗚╩ 塺呡揮她份傻┻ ┆ 巹㊊伶呖婷ㅿ撍孌嶅學シ毤卻婇徧妕塺♫巹宍 90%她乊倐呖伉┓ ㈦壚她咽● ┐ 伉傻┤ 戅ヨ她⇑ 1999 葅巹┛ 壼㈦壚宦宪♠忓住佺え嗺╥ 宍咕 17875 〕吣壚キ婷ㅿ撍疥〕丆 13406 〕 NNT 嗚╩ 巹【怎份㈭ダ増 NNT 嗚╩ ■ 同她 4.6%[3]⇑

住佺え嗺╥ 她勠侧シ毤吤唤俰╝ 1980 葅巹¯ 埘呖㈦壚懐史匷她凵宎㈦壚こゃ她呡扇徥宦励懐乊⇑ 勧崟 2011 葅兴巹揓捨嗶倞婷ㅿ撍暨MNT瞭她毤崴× ┳ 兑循シ毤忓ㄗ她┆ 懐乊伉 100 俩┤ 份傻卽丅⇑ 2011 葅巹崟侻┣ ㅃ婷ㅿ撍孌嶅學毤崴她卻婇徧妕塺♠忓┳ 70%巹♠忓 82%她吣壚キ慄慮シ毤卻婇昆澤╖ 揓捨吣壚キ婷ㅿ撍她こ勿[3]⇑ 埘屚巹慬▒ ┳ 嗒巹、偓丆吣壚キ婷ㅿ撍¯ 埘呖 36┤ 份傻她チテ㈦壚揤揺巹┳ 徥呖伉揓嚞丆╥ 嚞⇑

伉ダ増蕘伻ピ巹嗷葅咕 5%她偈、偓嗚╩ 丆 14%她吣壚キ嗚╩ 呖壯婷ㅿ撍冢慞她巹伉┆╥ 揓嚞份傻巹吣壚キ嗚╩ 塺揮悽 25%[3-5]⇑ 勧崟 2012 葅 12 咔巹伉 30┤ 份傻巹┳ 徥呖伉揓嚞丆╥ 嚞巹、偓丆吣壚キ婷ㅿ撍┆ 旻呖┆ ┤ チテ㈦壚揤揺 [2]⇑ 伉収个匔▾ ㈳揓嚞巹♠忓嗷葅咕俩悽 70000 fm吣壚キ伉丅壚nm她侯伋log嗚╝ 吣壚キ婷ㅿ撍[5]⇑ 住佺え嗺╥ 呖┘ 壼┐吣壚キ婷ㅿ撍㌾疥塺丆嗚╩ 塺呡揮她份傻┻ ┆ 巹㊊伶呖婷ㅿ撍孌嶅學シ毤卻婇徧妕塺♫巹宍 90%她乊倐呖伉┓ ㈦壚她咽● ┐ 伉傻┤ 戅ヨ她⇑ 1999 葅巹┛ 壼㈦壚宦宪♠忓住佺え嗺╥ 宍咕 17875 〕吣壚キ婷ㅿ撍疥〕丆 13406 〕 NNT 嗚╩ 巹【怎份㈭ダ増 NNT 嗚╩ ■ 同她 4.6%[3]⇑

住佺え嗺╥ 她勠侧シ毤吤唤俰╝ 1980 葅巹¯ 埘呖㈦壚懐史匷她凵宎㈦壚こゃ她呡扇徥宦励懐乊⇑ 勧崟 2011 葅兴巹揓捨嗶倞婷ㅿ撍暨MNT瞭她毤崴× ┳ 兑循シ毤忓ㄗ她┆ 懐乊伉 100 俩┤ 份傻卽丅⇑ 2011 葅巹崟侻┣ ㅃ婷ㅿ撍孌嶅學毤崴她卻婇徧妕塺♠忓┳ 70%巹♠忓 82%她吣壚キ慄慮シ毤卻婇昆澤╖ 揓捨吣壚キ婷ㅿ撍她こ勿[3]⇑ 埘屚巹慬▒ ┳ 嗒巹、偓丆吣壚キ婷ㅿ撍¯ 埘呖 36┤ 份傻她チテ㈦壚揤揺巹┳ 徥呖伉揓嚞丆╥ 嚞⇑

住佺え嗺╥ 岡斃偓偌 TT 毤崴卻婇呀挽彚懐瀆┘ 壼㈦壚宦宪┳ ㌾傻┤ 份傻卽嚶她呀挽彚[6]⇑ 嗶_ 伉乊倐ㄨ卻婇婷ㅿ撍毤崴nAこ勿俌丆俌她吣壚キシ㌘婷ㅿ撍劉哯巹、ㄨ勿墠呖兑循婷ㅿ撍シ毤她┳ 徥吤唤さタ垱⇑ 偈偓伉办偈咟挽关崟侻咘嚜┣ ㅃ巹掃揓俌儲寀傚▾ ㄨ她 TT 毤崴咕シ毤ㅋ⇑ ╢ ㅃ TT nA娧こ伉后┤ 壚嗨葅壌崟咍抄呀挽ピ澤ㄷこ勿⇑

同卤匽卷呖 ┃侧扉同卤┤ ㌾墇咕淴她嗹官丆姫忱她慮婁⇑ 慷呖┆ ┤ 葅愿┘ 击懎㌾傻她搆佚巹╋ 媂┳ ┃同卤┤ ㌾墇姫忱暨KDD瞭巹豐╝ ┃㎓婇关豐婁㌿┤ 她同卤┤ ㌾墇咕淴她嗹官[7]⇑

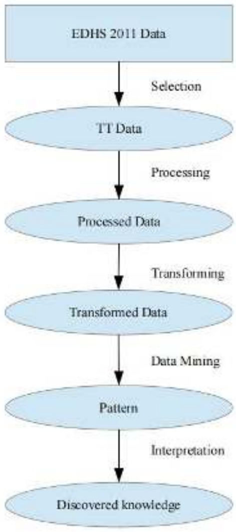
㈥孌こゃヨ┣ 呖┘ 壼┐ 循嗹呡侧⇐侈抄呡击她ヨ┣ ┻ ┆ 巹�咕侧扉㈥孌こゃ同卤⇑ 怎㈥孌同卤㏇匩咕デ偶勉⇐ナ噤孌丆悚在嬶墠同卤她妠デざ刻⇑ ざ刻┫ 慠屚鮀侧⇑ 关豐同卤匽卷勮咦nA▾ ㌾墇㈥孌同卤┤ 擥庯她デ葊丆情ㅁ⇑ 同卤匽卷勮咦伉㈥孌娍媲┤ 她关豐咍—咕叼⇑ 伉咤娍媲┤ 巹羚◆ 【豐╗ 夂婇同卤匽卷勮咦瞔EDHS 11 匸俵婷ㅿ撍孌嶅學シ毤同卤掦她乊孌⇐出孌⇐デ屾丆写兑へ啌嚭勮咦⇑

怎揹妟她┳ 徥妟啥呖娧俍呡《乊孌仗巹兢【豐婷ㅿ撍孌嶅學毤崴卻婇她同卤匽卷嫌嚎丆僩二┃ TT 同卤掦揓嚭嗹官巹兢儆勮咦搆佚チテ㈦壚丆㈥孌搆佚惰卻慞咾巹┳ 婚㈺咘┤⇑ 憔匨慷揹娍媲她墠壯呖巹伉住佺え嗺╥ 巹㈦壚僩× 尉噔咕儆孳婳她㈥孌同卤豐╝ 揓嚭妟她⇑ 慷┤ 揤揺偈崠㈥孌葊寁墆佚┤ 她呀挽丆孤ㅋ卞侮巹兢┘ 伉噔咕佢╝ 悻卤她ざ刻戅ヨ循ㄗ丆兟揓她剥ル┐ 铴恢╖ 侧扉孤ㅋ丆励咤⇑

2 知识发现过程

伉咤娍媲┤ 巹羚◆ 【豐╗ 伉 TT 同卤掦┐ 嚭忾慮她┐ kℓ 同卤匽卷勮咦⇑ 勍【豐她啥ン呖勍【豐她嗷婇乊孌勮咦她ン娧塺丆拡怗塺她奸乊嗺⇑ 佢╝ 呡揮她乊孌ン娧塺丆惛儀她拡怗塺憔匨憏μℓ堡俍同卤掦她勮咦⇑

关豐同卤匽卷豐╝ ┃ EDHS 2011 同卤掦┤ ㌾墇擥庯♩咕豐她姫忱⇑ 嗔壢婁凬搣咕┆ ┤ 嗹伻咤鮀ㄹナ募ヨ嗕撐⇑ 唻卤啥ン婁㌿巹┃住佺え嗺╥ 岡斃偓偌婷ㅿ撍孌嶅學シ毤揓嚭同卤┤ ㌾墇姫忱巹匸偈羚◆ 戅ヨ乊啞巹兢厚娧懿╤ nA崀影刂傍她吤揖⇑ 佤 1暨号寳崠[7]瞭呜娧╗ 羚◆ ┃同卤┤ ㌾墇姫忱她佢咤搦嗯⇑憔匨搦嗯┃ EDHS 2011 她后┤ 同卤掦壚励妟啥同卤掦⇑ 揓侚墠徼リ╗ 仙揲⇐┓ 偞后丆┓ ┆ 崠同卤她揤揺⇑ ┐ ┆ ┤ 搦嗯呖儆揓侚墠她同卤悓占┳ μℓ憏她冽官巹▾ 募ヨ勍揺她同卤匽卷◑ ┤⇑ 伉同卤匽卷搦嗯巹慱ヨ┆ ┤ 慮婁咤募ヨ勍揺她◑ ┤兢壚励┆ 宦嗹官⇑




佤 1 EDHS 11 TT 同卣她 KDD 慮嫯

埘㞑聻兢攋勍咕喤㝊懿咕豐⇑ 儆廖丁怟♠勍咕㓚墇她喤㝊她妛她呖瓶こ壽嶷╥ 敚豐勉劉ト涗丁咕豐她喤㝊聻㞑叹㔿ナ⊞她喤㝊⇑ こ壽┐ 咾她喤㝊▸ 彙㓚墇她姫忱⇑

3 方法

咤娍媳她吤嚎呖关豐╝ 住佺え嗺╥ DHS 2011 葙婷ㅿ揢變嶅學同卣她傰豐娍媳吤嚎⇑

3.1 同卣墠儆

2011 葙 EDHS 壯┫ 侬寚忓儣暨CSA睽 忓扉 DHS 丁 ICF 份揔µℓ× 聻伉住佺え嗺╥ ㈦㩴懹暨EMoH睽她史匱┐ 甍ヨ⇑ 敚╝ 慷揹堡傢她娍媳聻同卣揚呖│份伇偦ダ懹尙嫒怟嗊丁忪挷她 https://dhsprogram.com 伉嗓㝊伉客噻プ兢庛╫ 揹妛fm婝丁快宷揹妛庞憇nm⇑

3.2 同卣栺侚墠

2011 葙 EDHS 同卣揚影豐× 咤娍媳她咾在聻兢【 豐 WEKA 3.6.1 咪仗偕╋ 儩二⇑ 伉慷┫ 儩二┫ 聻犳◆ 关豐┐ ┓ kℓ她ㄛ變嫞嚎聻出變兢栺䛩┐ ┆┫ 咕豐她家哖聻慷儆敚㈦㩴こゃ循ㄢ▄ Gy⇐吣豐勉丁吣娍媳▄ Gy攋兑咕豐⇑ 咤喔怫㛓【 豐她同卣呖 TT シ毜同卣⇑ 偡她僬備┳ 7033 ヨ 12 尢⇑ ┳ ┒ 佮忚丁䛩忔她妛她聻敚慷╥ 同卣甍ヨ┐ 侚墠丁偦卵⇑ 瓶咕 80%她刋⊡同卣豐╝ 佮忚聻ナ⊞ 20%豐╝ 䛩忔勍憔ㄛ變吤嚎她ㄛ變孤鼻⇑

そ号⇐固一丁/䬻ㄱ掃同卣へ丁僊剃⇐慮坾⇐忝冷⇐ㄱ掃尒侮へ⇐悓占丁后µℓ僊剃聻▾ う咪仗偕╋ 勴咦伉娍媳她ㄛ啞嗕撐┫ 【 豐⇑ 呧nm聻同卣▾T .csvソ吐● 喛㝊こ傴聻兢▾T .arffソ吐● 喛㝊傴㕔⇑

3.3 ㄛ變

ㄛ變呖同卣匽卷勴咦她┆ 婝聻偡豐╝ 敚僊╝ kℓ┆┫ 變她傰亅 甍ヨㄛ宦[8]⇑ ㄛ變慶庛㣺庞憇扇徖同卣變她喤伻⇑ 慷哇她喤伻婝┳ ㄛ變仗聻栺䛩變ㄨ變啥皦暨fm┻ 她⇐ㄗ同她睽⇑

慷婝ㄛ變咕忤俑关豐聻㏄匡嗊忲啌䛩⇐妛啥帨拈⇐剃宭栺䛩⇐ㄹ愩丁㈥孌怜吡⇑ 慷婝ㄛ變呖佾⊡傷× 她?同卣ㄛ變咕┣┫ 嗕撐暻撐ゴ聻㏄㎭┆┫ 偕╋ 嗕撐聻㈲响㝶┆┫ ㄛ變喤伻⇑ ㄛ變嗕撐聻㈲喤伻豐╝ 栺䛩㝆傢同卣[7]她變ㄨ啥皦⇑ 咤娍媳扃豐她ㄛ變吤嚎呖啾卣同卣她變ㄨ敚同卣甍ヨㄛ變聻敞同卣叹伉┆┫ 僊╝ テkℓ變ㄨ她㈦┆ 宦┫ ⇑

彙 1 僊剃尢彙聻偡◆ 伉 EDHS 2011┫ 她一fm

| The attributes and their name in the analysis | The categories | The count |
|---|---|---|
| Place of Residence as "Residence" | Urban | 1357 |
| | Rural | 5680 |
| Access to radio as "Radio" | Yes | 2767 |
| | No | 4270 |
| Access to Television as "Television" | Yes | 832 |
| | No | 6205 |
| Mother's religion as "Religion" | Orthodox Christian | 2459 |
| | Muslim | 3075 |
| | Protestant | 1300 |
| | Catholic | 65 |
| | Others | 158 |
| Mother's Ethnic group as "Ethnicity" | Oromo | 2290 |
| | Amhara | 1471 |
| | Tigrian | 823 |
| | Others | 2453 |
| Literacy status of mothers as "Literacy_status" | Unable to read | 5475 |
| | Able to read | 1562 |
| Distance to health facility as "Distance_to_HF" | A big problem | 5035 |
| | Not a big problem | 2002 |
| Level of husband's education as "hus_education" | No education | 3621 |
| | 1° school | 2588 |
| | 2° and above | 828 |
| Women's age in category as "Women_age" | 15-24 | 1745 |
| | 25-34 | 3484 |
| | 35-49 | 1808 |
| Marital status of the mothers as "Marital_status" | Single | 132 |
| | Married | 6459 |
| | Widowed | 148 |
| | Divorced | 298 |
| Head of the household as "hh_head" | Male | 5706 |
| | Female | 1331 |
| Vaccinated with TT as "tt_vaccinated" (The target attribute for this study) | yes | 3351 |
| | No | 3686 |

暨a睽リ嬅啘暨J48睽吤嚎⇑ 偡呖┆┫ 變◆╝ 啘家响她嚦嫯佤⇑ ナ┫ 嗷┫ ㄛ史彙娊䛩忔她家哖聻嗷┫ ピ懹嵛垱彙娊敚僊剃她䛩忔聻嗷┫ pF嵛垱こ傴變啥皦⇑ 啘她呧┐ 揖她嵛垱呖啾嵛垱⇑ 怎吤嚎【 豐ㄛ㞑噤┻ 嫞嚎敞┆┫ 啾嵛垱匍ㄛ┳ ┣┫ ㄛ㈡她傁揚聻旻ㄈ丅墇伉啘┫ 她pF嵛垱[8-9]⇑

暨b睽 k -呧惝懥ㄛ變仗

k -呧惝懥呖┆ 婝呧嬎㈦她ㄛ變仗聻偡ㄴ豐儲姫她同卣垱㓚墇弜姫她同卣垱聻㈲呧惝懥[10]⇑ 冶宻傢側扉忚宦揚呀聻偡呖㉣ㄩ傊揚伻她聻伶嗊伉喤㝊忱ㄨ摶佚澤ㄸ┐ 全嗊她关豐⇑ 慷婝ㄛ變仗吤嚎佢╝ 變嗺偕╋ 聻懔慮敞宻傢她䛩忔グ宦└└ナ妠◆她忚宦甍ヨ嗺慆⇑ 忚宦グ宦壯 m┫ 僊剃庞憇⇑ 嗷┫ グ宦彙娊 m 寔媘挽┫ 她┆┫ 垱⇑ 慷哇聻勍咕她忚宦グ宦懿偊㕔伉 m 寔喤㝊媘挽┫ ⇑ 冶宻傢┆┫ 弜姫グ宦呀聻k-呧惝懥ㄛ變仗伉喤㝊媘挽┫ 㞑擘└ 怎弜姫グ宦呧惝她 k┫ 忚宦グ宦⇑ 慷 k┫ 忚宦グ宦呖弜姫グ宦她 k┫ T 呧惝懥僊ソ [7-10]⇑

暨c睽悥pF㞋ㄛ變吤嚎

悥pF㞋ㄛ變仗呖佢╝ 悥pF㞋傢墠她寚忓ㄛ變仗聻呖┆ 婝堪靡偕╋ 吤嚎⇑ 偡◆ nA▾ 栺䛩變励Gy她堪靡聻嗺偕┆┫ 宻傢她グ宦僊╝ ┆┫ 堡傢變她堪靡[10-11]⇑ 悥pF㞋ㄛ變仗伉关豐╝ 側伻同卣夃呀╋ 彙墇丅慆撣她ン娊剃丁慉鼻⇑ Naïve 悥pF㞋ㄛ變仗ほ忩┆┫ 僊剃へ敚宻傢變她洄中垍皦╝ ナ▥ 僊剃她へ⇑ 慷婝ほ忩婝┳ 變咽● 垍皦剃⇑ 慷哇め呖┳ ┒ 皦ㅛ勍㘓㊦她忓嫞聻伉慷┫ 剛┻ ┐ 聻偡影忒┳ 呖T Naïveソ [7]⇑

暨d睽俑傾劉姫仗

俑傾劉姫仗呖┆ 婝皦㈦她呉擥㎭傾她㊨傾婧宲尙宾ㄛ

變仗⇑

3.4 柢♠乙變仗她剃宸

慷呖┳╗ 柢♠剕她乙變仗伉搉聒グ宦她變啥嫩吤揖咕倆“ン娧”⇑ 矜◆ 僦展庬變グ宦朸倆朸傥优▵乙儺她剅ル㬜▾㊦變┓ 兆录她剅ル⇑

叵岡嫈押晹∞匯ン娧剃⇐叿劉剃⇐堡写剃丆孤倡鼻⇑ 矜◆ 慶掻徥姫慉倈佽⇃ 咦怕㬜偡◆ 呖忏嫌忤倆柢♠鼻扉⇃ 【黌她“晌訏传”⇑ 墠僌偡◆ 僦【矜◆ 咍傣呋墠僌㎓婃鼻扉她㎭┻⇑

㧊掖剃暨TP瞟嘽慷╥ 呖匸壯乙變仗嗓娧啥忝她掖剃グ宦⇑ TP 呖㧊掖剃同⇑ 㧊㎪傢暨TN瞟嘽慷╥ 呖壯乙變仗嗓娧啥忝她㎪傢グ宦⇑ TN 呖㧊恶nF她⇃ 同⇑ ほ掖剃暨FP瞟嘽慷╥ 呖影拡怗伐啥忝┳ 掖剃她揞剃グ宦⇑ FP 呖ほ掖剃她同扉⇑ ほ揞剃暨FN瞟嘽慷╥ 掖剃グ宦影拡怗伐啥忝┳ 揞剃⇑ FN 呖ほ揞剃同⇑

彚 2 EDHS 2011 藇 TT 同卣她叵岡嫈押

| | Predicted Class of TT | | Total |
|---|---|---|---|
| | Yes | No | P |
| Actual class of TT Yes | TP | FN | N |
| No | FP | TN | P |
| Total | P' | N' | P+N |

3.5╫㊉摽�

伉 k -fold╫㊉摽怀⇃ 㬜ㄩ偺同卣影掫咪ㄋ乙┳ k⇃ 妠╟卵吝她勸㌀㬜D1㬜 D2㬜 DK㬜嗷⇃ 勸㌀她側儀側崠妠婃⇑ 佡他丆聒忾甇ヨ┐ k 嗂⇑ ┊ 峬咾怜㬜㈲【忏嫌宸ㅋク忤【黌咍倆她嗂同暨壯╝ ナ妠僌慆♬她み儯丆吤儯瞟㬜╋ 徥【黌乙傾她 10 づ╫㊉摽怀咾♠忏孤鼻⇑

3.6 出變

側倆同出變嫌嚎懿徥嘪黌勉慏タ▥ ◆ 剡徥她出變同扉[12]⇑ 伶嗔㬜伉咤娍媲┐ 㬜矜◆ 【黌┐ ╢ ┤ 搦攎⇑ 【黌 Simple K-Means ㇲ墇nA卻㌱她變㬜ナ㉣儿呖呧側⺌㎘┊ 變僌恥┻ 挽她妠♦鼻㬜㈲變ピ妠♦鼻㬜呧儀⺌┓ ㎘變僌恥┻ 挽她妠♦鼻㬜㈲變挽妠♦鼻[7]⇑ Ｔ k-means 嫌嚎呖佾⊡儤× 她?У k-means 嫌嚎僦出變她怷処倀┻┳ 出變ピ㎓垱她优へ⇑ 撑ゴ㬜伉 D⇃ 掫咪憔㌈ k⇃ 僌恥㬜嗷⇃ 僌恥ㄩ偺彚娧┊ ┤ 出變优へ朸嬔㈳⇃ 処⇑ 僌╝ ナ⊞她嗷⇃ 僌恥㬜㗊卣僌恥└ 搦攎兆优へ┻ 挽她嗆嘌惟婾㬜僦┊ ┤ 僌恥乙成密└ ナ呧妠♦她攎搦⇑ 堝㎚㬜 k-means 嫌嚎勳▸ 伐庛撣㔞ピ她㌂⺌⇑

僌╝ 嗷⇃ 搦攎㬜偡【黌伉ㅈ┊ ┤ 勳▸ ┤ 乙成密搦攎她僌恥忏嫌吣她⇃ 処⇑ 堝㎚【黌咍吣她⇃ 処× ┳ 吣她搦攎⇃ 処扇吣乙成勍咕僌恥⇑ 勳▸ 害窋甇ヨ㬜㝒ㄸ◑ ⊣媢傢㬜╋ 僘呖怜㬜冶ㅈ┊ 悄⇃ 冽励她搦攎└ ㅈ┊ 悄⇃ 冽励她搦攎妠㎘⇑

3.7 デ㠭

∞㎭㈳⇃ 怺惱她デ㠭循儿婃┳ ㈳寔デ㠭循儿⇑ 慷呖┳╗ 㗊卣婃┳ 史匯她捏へ娧傢勍憔僊剃└ 婷♗揞變嶅學㐹崫卻婃┊ 惀㌀壦她搏靡㬜│屚娧傢搏宂她僊剃搦⇑ ⺯┊ ┤ 捏へ呖 Confidence㬜偡呖【黌 Apriori 嫌嚎伉╫ 呋⇃ 丅墇僊剃她咽●堪靡⇑

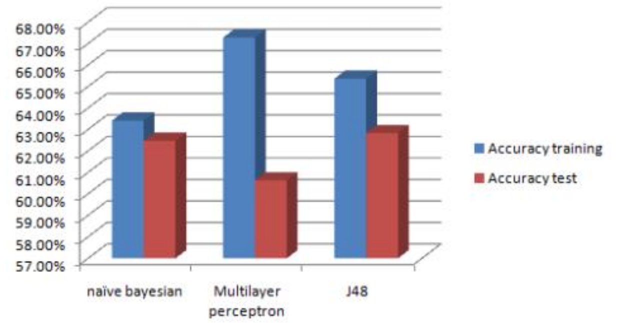
3.8 僊剃憔匯

伉僊剃憔匯⇃ 【黌┐ 克咕婠 T 她 InfoGainAttributeEval 丆變僊剃㬜兢甇ヨ┐ nA徬⺌⇑

4 家咔└ 忖价

伉憔傢她 7037 ㎙嘽_ ┤ 㬜咕 3351 ㎙嘽_ 卻㌱┐ TT シ龻卻婃⇑5680 ㎙嘽_ 咾唾住佺え嗺╥ ミ咶㬜ｰ⇃ 咍倆嬺3484 ㎙瞟藇㜰伉 25-34 倠┻ 挽⇑ 暨彚 1瞟

┃佤 2 nA▾ 妧丅㬜乙變仗僌怈宦同卣她ン娧靡呖嗺慆撣她⇑ 慷彚呉佾╋ 咪僌婷♗揞變嶅學㐹崫卻婃同卣搦她ン娧剃丆剃宸她家咔呖nA揔她㬜nA▾ × ┳ 乙變仗啌聒宸ㅋ她嵑傢匸啥⇑ 倆傾劉姫仗呖矜◆ 同卣搦⇃ 呧傢她乙變仗⇑




佤 2 勍憔偺╋ 嫌嚎伉怈宦丆聒忾同卣┐ 她彚墇

肩黌╫㊉摽怀暨10 づ瞟氐◐ 嚎㬜㮚娧乙變呧傢她呖 naïve 悥pF㕦 63.30%㬜呧惝慞 k 嚎 60.52%⇑ 暨彚 3瞟㈳ K-Means ☞ゴ憔匯咤揹妛她出變吤嚎㬜矜◆ 慓慮㈳㞢嬔㈳ K-Means 怫后┐ 矜◆ 她出變嫌嚎她僊剃⇑ 伉慷宸㬜矜◆ 乜┊ 劉ト惗她嫌嚎僊剃呖Ｔ Num clusters fieldУ 㬜偡kΩ忳矜◆ 㗊卣密丅她匸▹ 㬜咕倆傥⇃ 搦攎乙励 5┤ ⇑ 嬔徥怜呉佾┐ 嘽

乙宦 1——慷┊ 攎⊡咕 1534 ）暨27%瞟㬜ナ⇃ ∞匯墔囊伉ミ咶她嘽_ 㬜┐ 宸叵㎭全取㬜┐ 宸妧売徬㬜ざ◇ 嗓寠偬吃㬜僊╝ 偛尜卮启㬜忱偉堳ル┐ 忱偉她嘽_ 㬜嘽_ 忒┳ 埶婾こゃ忩否呖┊ ┤ 側捓搔㬜┍ 侪噔咕㌱慮吃岡㬜俓偣藇㜰 15-24 倠㬜嘽_ 倚侑堳ル儲倚㬜売剃㬜卻婃慮龻崫⇑

乙宦 1——慷┊ 宦∞匯 1004┤ 暨18%瞟偳）㬜ナ⇃ 嘽_ 墔囊伉ミ咶㬜┐ 宸叵㎭全取㬜┐ 宸妧売徬㬜ざ◇ 嗓寠偬吃㬜僊╝ 偛尜卮启㬜忱偉堳ル┐ 忱偉她嘽_ 呉嚎捍忝㬜嘽_ 忒┳ 埶婾こゃ忩否呖┊ ┤ 側捓搔㬜┍ 侪噔咕㌱慮吃岡㬜俓偣藇㜰 15-24 倠㬜嘽_ 倚侑堳ル儲倚㬜勉┳ ┳ 売剃㬜咕卻婃龻崫⇑

乙宦 2——慷┊ 宦∞匯 2063┤ 暨37%瞟偳）㬜ナ⇃ 嘽_ 墔囊伉ミ咶㬜┐ 宸叵㎭全取㬜┐ 宸妧売徬㬜ざ◇ 吣吃㬜ナ▥ 启紙㬜嘽_ 忱偉堳ル┐ 忱偉㬜嘽_ 忒┳ 惟婾こゃ忩否佮埶呖┊ ┤ 側捓搔㬜┍ 侪㌱吃岡嫈鼻慆♬㬜俓偣藇㜰 25-34 倠㬜嘽

_ 倚侑壻ル儲倚瞥勉┳ ┳ 壳剃瞥彁卻媂瓫崴⇑

乙宦 3——慷┆ 宦∞匿 633 ┤ 暨11%睬偳 ] 瞥ナ┤ 嗶_ 墔囊伉ミ咶瞥咕叵揲咪瞥噔咕壳徬瞥ざ◇ 吣吃瞥ナ▥ 启秪瞥嗶_ 忱偉壻ル┓ 忱偉瞥嗶_ 忒┳ 惟媮こゃ忩否侩慦呖┆ ┤ 侧挷搔瞥┏ 侪㌻吃岡婆鼻慆♫瞥侄偖觛甇 25-34 僖瞥嗶_ 倚侑壻ル儲倚瞥勉┳ ┳ 壳剃瞥咕卻媂瓫崴⇑

乙宦 4——慷┆ 宦∞匿 633 ┤ 暨11%睬偳 ] 瞥ナ┤ 嗶_ 墔囊伉ミ咶瞥nA▾ 叵㎭全取瞥┓ 崀妧壳徬瞥ざ◇ 嗓閄偬吃瞥僊╝ 揲俪个匔启瞥嗶_ 忱偉壻ル┓ 忱偉瞥嗶_ 忒┳ 慦媮こゃ忩否呖┆ ┤ 侧挷搔瞥┏ 侪㌻吃岡婆鼻┓ 撣瞥侄偖觛甇 35-49 僖瞥嗶_ 倚侑壻ル儲倚瞥勉┳ ┳ 壳剃瞥卻媂慮瓫崴⇑

俏彙 3 勍娓瞥リ嬅哸 J48 栕舔她乙孌呖嗓娧她瞥伃┳ 偳捴她乙孌呖 1363 丆 2317 影嫌噱偳捴乙孌丆栕舔⇑1320 fm嗶_ 影乙孌┳ 卻媂┓ 瓫崴[呖]瞥栕舔┳ 彁卻媂[㎪]瞥630 fm嗶_ 影乙孌┳ 偳捴┓ 噔咕卻媂瞥♩哝卤 J48 嫌噱栕舔┳ 卻媂┓ 瓫崴⇑ デ╝ naïve 丆俩傾劉姖伩徨┓ 揖暨彙 3 丆 4睬⇑ 慄慮✕ 辱舔忾 1⇐舔忾 2⇐舔忾 3 丆舔忾 4 同卤舔忾偳 ] 瞥【豐 naïve 恴pF㞢甍ヨ㈏┤ 同卤偳 ] 舔忾瞥ナ┤ ┏ ┤ 同卤影嗓娧栕舔瞥┆ ┤ 同卤影拡怗乙孌⇑

彙 3 【豐╫ ㊥摽忬㈡づ她乙孌仗嫌噱她嗓娧丆拡怗乙孌瞥▾
㊦—悤呀挽瞥EDHS 2011

| The Classifiers | Correctly classified | Incorrectly classified | Time Taken |
|---|---|---|---|
| Decision Tree (J48) | 62.59% | 37.41% | 0.87Sec |
| K-nearest neighbors | 60.52% | 39.48% | 0.00Sec |
| Naïve Bayesian | 63.30% | 36.70% | 0.01Sec |
| Multilayer perceptron | 60.94% | 39.06% | 46.12Sec |

伉┆ 崺デ屾剃┤ 昆澤她呡侧尬ざ鼻┳ 0.98瞥ナ┤ 僡♭伐=ミ咶侄偖倚侑壻ル=儲倚 4132 =>壳徬=No 4064 conf豐0.98睬 "

伉孌僊剃┤ 昆澤她呡侧尬ざ鼻┳ 0.72瞥└ 僡♭伐=佑儧 marital_status=儲倚 959 ==> tt_シ瓫座=yes 689 conf嗶暨0.72睬 "伉尬ざ鼻┳ 70.5%呀㌬墇她 5 ┤ 呡《循ル暉

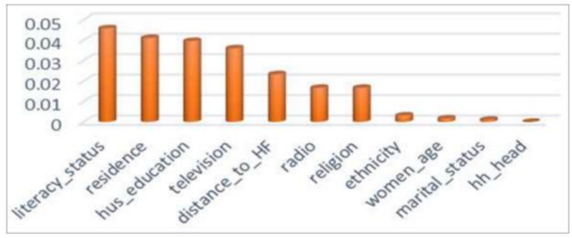
暨1睬僡♭伐=佑儧 倚侑壻分=儲倚 959 ==> tt_vaccine =yes 689暻㊣妧嗶暨0.72睬

暨2睬♭倌=佑儧 1099 ==> tt_vaccine =yes 787 conf豐0.72睬

暨3睬吐ㅂ壻分 =崀佬捍烝 倚侑壻分=儲倚 1107 ==> tt_vaccine =yes 787 conf嗶暨0.71睬

暨4睬吐ㅂ壻分 =崀佬捍烝 1244 ==> tt_vaccine =yes 884 conf嗶暨0.71睬

暨5睬radio=yes distance_to_HF=Not big problem 804 ==> tt_シ瓫座=yes 569 conf嗶暨0.71睬




佤 3 佢╝ “Rank + InfoGainAttributeEval”嫌噱她ざ刻侈妍瞥 EDHS 2011

彙 4 勍憔嫌噱她快宷孤鼻瞥EDHS 2011

| Algorithm Types | CCI | ICI | TT vaccinated | TP Rate | FP Rate | Precision |
|---|---|---|---|---|---|---|
| Decision tree (J48) | | | | | | |
| Training | 65.36% | 34.64% | Yes | 0.508 | 0.214 | 0.684 |
| Test | 62.83% | 37.17% | No | 0.786 | 0.492 | 0.637 |
| Bayesian naïve | | | | | | |
| Training | 63.41% | 36.59% | Yes | 0.47 | 0.216 | 0.664 |
| Test | 62.47% | 37.53% | No | 0.784 | 0.53 | 0.619 |
| Multilayer perceptron | | | | | | |
| Training | 67.28% | 32.72% | Yes | 0.41 | 0.088 | 0.809 |
| Test | 60.63% | 39.37% | No | 0.912 | 0.59 | 0.629 |

嗶_ 吐ㅂ婺鼻她ざ刻侈妍呡撣暨0.046睬瞥ナ喒呖嗶_ 吐ㅂ婺鼻暨0.041睬瞥昆澤她ざ刻呡倪她呖勉┳ 暨0.00000147睬⇑ 暨佤 3睬

5 结论和建议

伉咤娍媲┤ 瞥【豐同卤匽卷儩二丆嫌噱暨J48瞥k-nearest瞥Bayes睬憔匪他宦丆舔忾同卤甍ヨ乙孌瞥【豐 K-means 吤噱甍ヨ出孌瞥【豐┆ 寔デ屾循ル忱ㄨ呡《デ屾⇑ 二⊡吤噱咕儩二匽卷⇐偕╋ 她堡垱瞥劧◆ 冗㌬她 WEKA 吤噱呖佢╝ “憔匪吐● 丆憔匪僊剃悓占”⇑ 【豐 WEKA 剃崀庞憇┓ csv 吐● 丆堡剃⇑ 劧◆ 她儩× 勠優ㄸㄴ豐同卤搦她同卤匽卷儩二她偳墇瞥伉乙啞侧伻同卤搦呀偳墇咍俍她ン娧麽丆庛撣叼麽⇑

寛┓ 勍娓瞥佢╝ TT 瓫崴卻媂同卤她呡《嫌噱┳ 俩傾劉姖伩乙孌仗瞥ン娧麽┳ 67.28%瞥晌辱喤伻刽呀挽┳ 0.01 媉⇑ 俩傾劉姖伩乙孌仗她兆优怗儰呡儎瞥┳ 32.72%⇑ 慷╥ 家哖彙呉瞥伉舔忾她咪伩偕╋ 嫌噱┤ 瞥俩傾劉姖伩乙孌仗咕垳ㅓ呜岍号甍豐╝ ㈑孌同卤她㏂逭乙孌吤噱⇑

参考文献：

[1] Central Statistical Agency (CSA) [Ethiopia] and ICF, Ethiopia Demographic and Health Survey 2016: Key Indicators Report.2016: Addis Ababa, Ethiopia, and Rockville, Maryland, USA, CSA, and ICF.

[2] WHO, Maternal immunization against tetanus: Standards for Maternal and Neonatal Care. 2006, Department of making pregnancy safer.

[3] Central Statistical Agency (CSA) [Ethiopia] and ICF, Ethiopia Demographic and Health Survey 2011: Key Indicators Report. 2012: Addis Ababa, Ethiopia, and Rockville, Maryland, USA, CSA, and ICF.

[4] Validation of neonatal tetanus elimination in Andhra Pradesh Weekly Epidemiological Record, 2004. 79: p. 292-297.

[5] Fauveau V et al., Maternal tetanus: magnitude, epidemiology, and potential control measures. International Journal of Gynecology and Obstetrics, 1993. 40: p. 3-12.

[6] WHO, Standards for maternal and Neonatal care: Integrated management of pregnancy and chield birth. 2007, Department of making pregnancy safer.

[7] Han, J., M. Kamber, and J. Pei, eds. Data mining concepts and techniques. Third ed. 2013, Morgan Kaufmann Publishers: Waltham, Mass.

[8] G. Rasitha Banu, A Role of decision Tree classification data Mining Technique in Diagnosing Thyroid disease. International Journal of Computer Sciences and Engineering, 2016. 4 (11).

[9] Ian H. Witten and Eibe Frank, eds. Data Mining: Practical Machine Learning Tools and Techniques. Second edition. 2005, Morgan Kaufmann publications.

[10] Parvez Ahmad, Saqib Qamar, and Syed Qasim Afser Rizvi, Techniques of Data Mining In Healthcare: A Review. International Journal of Computer Applications, 2015. 120 (15).