REFERENCES

Afzal, H. & K.Mehmood. 2016. "Spam Filtering of Bi-Lingual Tweets Using Machine Learning".*Conference: 2016 18th International Conference on Advanced Communication Technology (ICACT)*.p. 714-718.

Ahmad, LG., AT.Eshlaghy.,A.Poorebrahmi., M.Ebrahimi & AR. Razavi. 2013. "Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence". *Health and Medical Informatics*. Vol. 4:124.

Ali,R., U. Ghani & A. Saeed.n.d."Data clustering and its applications".*TRIPOD* <http://members.tripod.com/asim_saeed/paper.htm>.accessed: 8 July 2014.

Almeida, T.A., G. Hidalgo, J.M.,& Yamakami. 2011. " Contributions to the Study of SMS Spam Filtering: New Collection and Results". *Proceedings of the 2011 ACM Symposium on Document Engineering (DOCENG'11)*. p. 259-262.

Amayri,O., & N.Bouguila. 2010. "A study of spam filtering using support vector machines".*Journal Artificial Intelligence Review*.Vol.34. p. 73-108.

Amin, M.N.& M.A.Habib. 2015."Comparison of Different Classification techniques using WEKA for Hematological Data".*American Journal of Engineering Research (AJER)*.Vol .4, (3): p. 55-61.

Arkib. 11 February 2013. "Kegiatan penipuan SMS spam membimbangkan SKMM". *UtusanMalaysia*.<http://www.utusan.com.my/utusan/Dalam_Negeri/20130211/dn_21/Kegiatan-penipuan-SMS-spam-membimbangkan-SKMM.> accessed:15 March 2014.

Babur, I.H., J. Ahmad, B.Ahmad & M.Habib. 2015. " Analysis of DBSCAN Clustering Techniques on different datasets using WEKA Tool".*Sci.Int.(Lahore)*,Vol. 27, p.5087-5090.

Bailey, R. 4 March 2014. "Blood".*About.com Biology*.<http://biology.about.com/od/humananatomybiology/a/blood.htm >. accessed: 13 January 2014.

Bala,R., S. Sikka & J. Singh.2014."A comparative Analysis of clustering algorithms".*International Journal of Computer Algorithms*. Vol.100, (15): p.35-39.

Berek, C.& M. Ziegner. 1993. "The Maturation of the Immune Response". New Jersey: Humana press. Vol.14. P. 400-404

Bhan,N & D.Mehrotra. 2013. "Comparative Study of EM and k-Means clustering techniques in WEKA interface".*International Journal of Advance Technology and Engineering Research (IJATER)*.Vol .3, (4): p. 40-44.

Bing, Z., Y. Yiyu & L. Jigang. 2010. "A Three-Way Decision Approach to Email Spam Filtering". *AI'10 Proceedings of the 23rd Canadian conference on Advances in Artificial Intelligence.*Vol. 6085, p.28-39.

British English SMS Corpora. 2011. <http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/. > accessed: 07 August 2014.

Burnet, F. M.1959."The Clonal Selection Theory of Acquired Immunity".*Nashville :Vanderbilt University Press.*

Business. 29 April 2013. Chat app messaging overtakes SMS texts, Informa says. BBC website < http://www.bbc.com/news/business-22334338 > accessed: 17 August 2015.

Chakraborty,S. & B.Mondal . 2012. "Spam mail Filtering using Different Decision Tree Classifiers through Data Mining Approach – A Comparative Performance Analysis". *International Journal of Computer Applications.* Vol. 47. p. 26-31.

Chaudhari,B & M.Parikh. 2012. "A Comparative Study of clustering algorithms using WEKA tools". *International Journal of Application or Innovation in Engineering and Management (IJAIEM).*Vol.1, (2): p. 154-158.

Chaurasia, V. & S.Pal. 2014."A novel approach for breast cancer detection using data mining techniques". *International Journal of Innovative Research in Computer and Communication Engineering*. Vol.2, (1): p. 2456-2465.

Chen, C.L, S.Frank, Tseng & T. Liang. 2010. " An Integration of WordNet and fuzzy association rule mining for multi-label document clustering" *Data and Knowledge Engineering.*Vol.69, (11): p. 1208-1266.

Chhabra, P.,R. Wadhvani & S.Shukla. 2010. "Spam Filtering using Support Vector Machine (SVM*)". International Conference [ACCTA-2010]*.Vol.1,(2,3,4):p.166-171.

Christina, V., S. Karpagavalli & G.Suganya. 2010. "Email spam filtering using supervised machine learning techniques". *International Journal on Computer Science and Engineering (IJCSE).*Vol. 2, (9): p.3126-3129.

Cloudmark. 17 Apr 2013. "2013's First Quarter at a Glance". *Cloudmark Security* <http://blog.cloudmark.com/2013/04/17/2013s-first-quarter-at-a-glance/>.accessed: 26 June 2015.

Cocotas, A. 2013. "Chart of the day: Kids Send a Mind Blogging Number of Texts Every Month" *Business Insider*. < http://www.businessinsider.com/chart-of-the-day-number-of-texts-sent-2013-3?IR=T&>accessed: 5 March 2014.

Cook, N.11 September 2013."SMS spam more dangerous than email spam". *The M2 Computing* <http://www.m2blog.co.uk/sms-spam-more-dangerous-than-email-spam/cook.>accessed: 8 March 2014.

de Castro, L.N & F.J. von Zuben.1999. "Artificial Immune Systems : Part II – A Survey of Application".<ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/tr_dca/trdca0199.pdf >. accessed: 16 April 2014.

de Castro, L.N &J. Timmis. 2002. "Artificial Immune System: A New Computational Intelligence Paradigma ".Springer-Veralg. London.

Delany, S.J.,M.Buckley & D.Greene. 2012. "SMS Spam Filtering: Methods and Data". *Expert Systems with Application*.Vol.39, (10): p.9899-9908.

Derek Johnson (2013). Why is a text message only 160 characters?. *tatango website*. http://www.tatango.com/blog/why-is-a-text-message-only-160-characters/ . Accessed on 16 June 2014.

DIT SMS Spam Dataset. 2012. *Dublin Institute of Technology*. <http://www.dit.ie/computing/research/resources/smsdata/>. accessed :05 August 2014.

Donovan, F. 2013. "Two-thirds of mobile phone users get SMS spam".*FierceMobileIT*.<http://www.fiercemobileit.com/story/two-thirds-mobile-phone-users-get-sms-spam/2013-08-17.>accessed: 6 March 2014.

Forrest, S., A.S. Perelson, L. Allen & R.Cherukuri. 1994. "Self-Nonself Discrimination in a Computer".*In Proceedings of 1994 IEEE Symposium on Research in Security and Privacy*.p. 202-212.

Godara,S. & A.Verma. 2013. "Analysis of various clustering algorithms". *International Journal of Innovative technology and Exploring Engineering (IJITEE)*. Vol.3, (1): p.186-189.

Greensmith, J., A. Whitbrook & U. Aickelin. 2010. *Artificial Immune System* (2nd edition). UK: Springer.

Haughn,M., S.Gibilisco & M.Rouse. November 2014. Confidentiality, integrity and availability (CIA triad).*Whatls.com*.< http://whatis.techtarget.com/definition/Confidentiality-integrity-and-availability-CIA > accessed: 17 August 2015.

Idris, I. 2012. "Model and Algorithm in Artificial Immune System for Spam Detection".*International Journal of Artificial Intelligence and Applications (IJALA)*, Vol.3, (1): p.83-94.

Idris, I. & A. S. Muhammad. 2012. "An Improved AIS based E-mail Classification Technique for Spam Detection". *The Eight International Conference on eLearning for Knowledge-Based Society*, 23-24 February 2012, Thailand.

Ilic, M., P.Spalevic.,M.Veinovic. &W.S.Alatresh. 2016. "Students' success prediction using Weka tool". *INFOTECH-JAHORINA*. Vol.15. p. 684-688.

Iqbah, M., M.M. Abid, M.Ahmad & F.Khurshid. 2016. "Study on the Effectiveness of Spam Detection Technologies". *I.J. Information Technology and Computer Science.* p.11-21.

Jerne, N.K. 1974. "Towards a Network Theory of the Immune System".*AnnImmunol (Inst.Pasteur).*p. 373-389.

Joe, I &H.Shim. 2010. "An SMS spam filtering system using support vector machine". *Second International Conference, FGIT 2010*. Vol. 6485. p. 577-584.

Johnson, D. 2011. "Texting Statistics by Age 2010".*Tatango.*<http://www.tatango.com/blog/texting-statistics-by-age-2010/.> accessed: 4 March 2014.

Johnson, D. 13 August 2013. "SMS Gift card Scams Down, SMS Job Listing Scams Up". *Tatango*<http://www.tatango.com/blog/sms-gift-card-scams-down-sms-job-listing-scams-up/.> accessed: 5 March 2014.

Johnson, D.2013. "Why is a text message only 160 characters?".*Tatango*<http://www.tatango.com/blog/why-is-a-text-message-only-160-characters/>.accessed:16 June 2014.

Junaid, M.B., & M.Farooq. 2011. "Using Evolutionary learning Classifiers to do Mobile Spam (SMS) Filtering". *GECCO'11Proceedings of the 13th Annual Conference on Genetic And evolutionary Computation.*p.1795-1802.

Khan, S.& M.A.Peer. 2013. "Evaluation of knowledge extraction using various classification data mining techniques". *International Journal of advanced Research in computer science and software engineering.* Vol. 3,(6): p. 251-256.

Khare, P., K. Burse &A.Pandey. 2016 "Comparing Various Classification Techniques Through Weka for Ovarian Cancer". *International Journal of Engineering and Advanced Technology (IJEAT).* Vol.5. p. 20-23.

Kumar, A. & I. Chatterjee. 2016. "Data Mining: An experimental approach with WEKA on UCI Dataset". *International Journal of Computer Applications.* Vol.138. p. 23-28.

Levine, J.R. 2005. "Experience with Greylisting". *Proceedings of second conference on Email and Anti-Spam (CEAS 2005).*p. 1-2.

Mahmoud, T.M & A.M. Mahfouz. 2012. "SMS Spam Filtering Technique Based on Artificial Immune System".*IJCSI International Journal of Computer Science Issue.* Vol.9,(1): p. 589-597.

Matzinger, P. 1994. "Tolerance, danger, and the extended family".*Annual Review of Immunology.*Vol.12 p.991-1045.

Matzinger, P. 1998. "An innate sense of danger".*Seminars in Immunology.*Vol.10 p.399-415.

Matzinger, P. 2002. "The Danger Model: A Renewed Sense of Self". *Science.*Vol.296 p.301-305.

Mohd Al Qayum Azizi. 7 June 2010. "Polis tumpaskan sindiket penipuan SMS" *.mStar.*<http://mstar.com.my/cerita.asp?sec=mstar_jenayah&file=/2010/6/7/mstar_jenayah/20100607145625.> accessed: 15 March 2014.

Mosquare,A.,L.Aouad, S.Grzonkowski & D.Morss. 2014. "On detecting messaging abuse in short text messages using linguistic and behavioural patterns". *arXiv.*

Mujtaba, G.& M,Yasin. 2014. "SMS Spam Detection Using Simple Message Content Features". *Journal of Basic and Applied Scientific Research.*Vol. 4.p. 275-279.

n.a."Hierarchical Clustering intro"*FrontlineSolvers.*<http://www.solver.com/hierarchical-clustering-intro>accessed: 17 July 2014.

n.a. 14 February 2011. "Viral Attack" *.ASU-School of Life Science.* <http://askabiologist.asu.edu/macrophage.>.accessed: 1 March 2014.

n.a. 2012."5 worried it's spam?5 things to look for".*Canada's Anti-Spam Legislation.* <http://fightspam.gc.ca/eic/site/030.nsf/eng/h_00241.html> .accessed: 16 June 2014.

n.a. 21 October 2013."Basic Immune Inspired Algorithms".*The Online Home of Artificial Immune Systems.*<http://www.artificial-immune-systems.org/algorithms.shtml > accessed:5 February 2014.

Nanda, S.J. 2009. *Artificial Immune Systems: Principle, Algorithms and Applications.*(Master of Technology).National Institute of technology Rourkela.

Nazri, M.Z.A, S.M. Shamsuddin, A. A. bakar & S.Abdullah. 2009. "A Hybrid Approach for Learning Concept Hierachical from Malay Text Using GAHC and Immune Network". *ICARIS '09 Proceedings of the 8th International Conference on Artificial Immune System.*p. 315-328.

Nazri, M.Z.A, S.M.Shamsuddin & A. A.bakar. 2010. "Clonal Selection Algorithm for Learning Concept Hierarchical from Malay Text". *In proceeding of Rough Set and Knowledge technology-5th International Conference, RSKT 2010.* p. 15-17.

Nosrati, L.,& A.N.Pour. 2011. "Dynamic Concept Drift Detection for Spam Email Filtering". *Proceedings of ACEEE 2nd International Conference on Advances Information and Communication Technologies (ICT 2011).*Vol.2.p.124-126.

Pallavi & S.Godara. 2011. "A Comparative Performance Analysis of Clustering Algorithms". *International Journal of Engineering Research and Applications (IJERA).*Vol. 1, (3): p. 441-445.

Pandya, S & P.V. Virparia. 2013."Comparing the applications of various algorithms of classification techniques of data mining in an Indian university to uncover hidden patterns".*International Journal of advanced research in computer science and software engineering*.Vol .3, (5): p. 1023-1026.

Patel,D., R.Modi & K.Sarvakar. 2014. "A Comparative Study of Clustering Data Mining: Techniques and Research Challenges". *IJLTEMAS*.Vol.3, (9):p. 67-70.

Perez, M. n.d."Humoral Response".*Immune*.< http://immuneweb.xxmu.edu.cn/immunology/humoral.html >accessed : 20 March 2014.

Pour, A.N., R. Kholghi &B, Roudsari.2012. "Minimizing the Time of Spam Mail Detection by Relocating Filtering System to the Sender Mail Server".*International Journal of Network Security and Its Applications*. Vol.4. p. 53-62.

Purves, K.W., D. Sadava,G.H.Orians & C.Heller. 2010. *The Science of Biology*. 8th edition.USA: W.H. Freeman & Company.

Rahman, R.M & F. Afroz. 2013. "Comparison of Various Classification Techniques Using Different Data Mining Tools for Diabetes Diagnosis".*Journal of Software Engineering and Applications*.Vol.6, (3): p. 85-97.

Ramachandran,A., D.Dagon & N.Feamster. 2006. "Can DNS-Based Blacklists keep up with bots?".*CEAS 2006-Third Conference on Email and Anti-Spam* .

Richter, F. 2013. "5 billion people to use mobile phone by 2017". *The Statistics Portal*.<http://www.statista.com/chart/1517/worldwide-mobile-phone-users/>.accessed: 5 June 2015.

Rouse, M. 2007. "Reverse DNS (rDNS) definition". <http://searchnetworking.techtarget.com/definition/reverse-DNS.>accessed: 01 March 2015.

S. Prabha, K.Duraiswamy & M.Sharmila. 2014. "Analysis of Different Clustering Techniques in Data and Text Mining". *International Journal of Computer Science Engineering (IJSSE)*.Vol.3, (2): p.107-116.

Saman Pooja Mittal. 2014. "Comparison and Analysis of various Clustering method in data Mining on education dataset using the weka tool". *International Journal of Emerging Trends and technology in Computer Science (IJETTCS)*.Vol.3, (2): p. 240-244.

Saxena,P. & S.Lehri. 2013. "Analysis of Various clustering algorithms of data mining on health informatics". *International Journal of Computer and Communication technology*.Vol.4, (2): p. 108-112.

Sehgal,G.& K.al Garg .2014."Comparison of various Clustering Algorithms".*International Journal of computer science and information techniques*. Vol. 5, (3): p.3074-3076.

Shah, N.& S.Mahajan. 2012. "Document Clustering: A Detailed Review". *International Journal of Applied Information Systems (IJAIS)*.Vol.4, (5): p.30-38.

Shahi, T.B.,& A.Yadav. 2013. "Mobile SMS Spam Filtering for Nepali Text Using Naïve Bayesian and Support Vector machine". *International Journal of Intelligence Science*.Vol.4.p.24- 28.

Sharma, B.R & A. Paul. 2013. "Clustering Algorithms: Study and Performance Evaluation using WEKA". *International Journal of Current Engineering and Technology*.Vol.3, (3): p. 1094-1098.

Sharma,N.,A.Bajpai & R.Litoriya. 2012. "Comparison the Various Clustering algorithms of WEKA tools". *International Journal of Emerging Technology and Advanced Engineering*.Vol.2, (5): p. 73-80.

Sharma,S& V.Gupta. 2013."Punjabi Documents Clustering System".*Journal of Emerging Technologies in Web Intelligence*.Vol.5, (2): p. 171-187.

SMS Spam Corpus v.0.1. 2011. <http://www.esp.uem.es/jmgomez/smsspamcorpus/>. accessed: 20 January 2014

Sohn, D-N., J-T.Lee & H-C.Rim . 2009."The Contribution of Stylistic Information to Content-based Mobile Spam Filtering". *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*.p.321-324.

Stepney, S., R.Smith, J.Timmis & A.Tyrrell. 2004. "Towards a conceptual framework for artificial immune system". *In Proc. of the 3rd International Conference on Artificial Immune Systems (ICARIS)*.Vol.3239.p.53-64.

Sulaiman,N.F & M.Z.Jali. 2014. "Integrated Mobile SPAM Model using Artificial Immune System Algorithms". *Knowledge Management International Conference 2014 (KMICe 2014).*

Sulaiman,N.F & M.Z. Jali. 2014. "A New SMS Spam Detection Method Using Both Content-Based and Non Content-Based Features".*2015 2nd International Conference on Communication and Computer Engineering (ICOCOE'2015).*

Suruhanjaya Komunikasi dan Multimedia Malaysia. 2009. "Statistical Brief Number Nine Hand Phone User Survey 2009". *SKMM* <http://www.skmm.gov.my/skmmgovmy/files/attachments/HPUS-2009.pdf > .accessed : 22 April 2014.

Tamarkin, A.D. 2011. "Lymphocytes".*Anatomy and Physiology*.<http://faculty.stcc.edu/AandP/AP/AP2pages/Units21to23/immune/lymphocytes.htm.>. accessed: 23 February 2014.

Tan,H.,N.Goharian & M,Sherr. 2012. "$100,000 Prize Jackpot. Call Now! Identifying the Pertinent Features of SMS Spam".*SIGIR'12* .

Tang,N. & V.RaoVemuri. 2005. "An Artificial Immune System Approach to DocumentClustering". *SAC '05 Proceedings of the 2005 ACM Symposium on Applied Computing*. p. 918-922.

Taylor, T. 2013. "Immune and Lymphatic System".*InnerBody*.<http://www.innerbody.com/image/lympov.html.>.accessed: 2 March 2014.

Timmis, J. & T. Knight. 2002. *Artificial Immune System: Using the Immune System as Inspiration for Data Mining*. United Kingdom: Group Idea Publishing.

Todar, K. 2012. "Immune Defense against Bacterial Pathogens: Adaptive or Acquired Immunity". *Todar'sOnline textbook of bacteriology*.<http://textbookofbacteriology.net/adaptive_2.html.>.accessed: 15 march 2014.

Uysal, A.K., S.Gunal.,S.Ergin & E.S.Gunal. 2013."The impact of feature extraction and selection on SMS spam filtering"*Elektronik airelektroteehnika*. Vol .19,(5): p. 67-72.

Vaas, L. 26 March 2014. "Chinese cops nab 1530 mobile SMS spammers in raid on fake base stations". *Nakedsecurity*.<http://nakedsecurity.sophos.com/2014/03/26/chinese-cops-nab-1530-mobile-sms-spammers-in-raid-on-fake-base-stations/.> accessed: 4 April 2014.

Verma,M., M. Srivastava, N. Chack, A. K.Diswar, & N.Gupta. 2012. "A Comparative Study of various clustering algorithm in data mining". *International Journal of Engineering Research and Applications (IJERA)*. Vol.2,(3): p.1379-1384.

Victor.,W.Onomza, I.Idris, Abdullahi,M.B.Hahimi,Danladi & A.Isah. 2013. "A Negative Selection Algorithm based on Email Classification Techniques". *World of Computer science andInformation Technology Journal (WCSIT)*. Vol.3, (3): p. 56-59.

Wahbeh, A.H & M. Al-Kabi. 2012. "Comparative Assessment of the performance of three WEKA text classifiers applied to Arabic text". *ABHATHAL-YARMOUK:"Basic Sci. &Eng"*.Vol. 21, (1): p. 15-28.

Witten, I.H & E.Frank.2005. *Data Mining Practical Machine Learning Tools and Techniques*.2nd edition.Jim Gray. (ed). San Francisco: Morgan Kaufmann.

Xu, Q., E.W.Xiang & Q.Yang, .2012."SMS Spam Detection Using Non-Content Features". *IEEE Intelligent Systems*. Vol.27, (6): p.44-51.

Zainal, K, N.F.Sulaiman. & M.Z. Jali. 2015. "An Analysis of Various Algorithms For Text Spam Classification and Clustering Using Rapid Miner and Weka". *International Journal of Computer Science and Information Security*. Vol. 13, (3): p.66-74.

Zhu, Y.&Y.Tan.2011. "A Danger Theory Inspired Learning Model and Its Application Spam detection". *ICSI 2011*.p. 382-389.

Zimmerman, A.K. 8 February 2013. "Lymphatic Systems: facts, Functions and Diseases". *LiveScience*.<http://www.livescience.com/26983-lymphatic-system.html.>. accessed: 15 February 2014.

APPENDICES

Appendix A

List of dataset used for classification

1. FadhilahSpam Dataset

Please call our customer service representative on FREEPHONE 0808 145 4742 between 9am-11pm as you have WON a guaranteed £1000 cash or £5000 prize!

Are you unique enough? Find out from 30th August. www.areyouunique.co.uk

500 New Mobiles from 2004, MUST GO! Txt: NOKIA to No: 89545 & collect yours today!From ONLY £1 www.4-tc.biz 2optout 087187262701.50gbp/mtmsg18

Will u meet ur dream partner soon? Is ur career off 2 a flyng start? 2 find out free, txt HORO followed by ur star sign, e. g. HORO ARIES

Text & meet someone sexy today. U can find a date or even flirt its up to U. Join 4 just 10p. REPLY with NAME & AGE eg Sam 25. 18 -msgreed@thirtyeight pence

U 447801259231 have a secret admirer who is looking 2 make contact with U-find out who they R*reveal who thinks UR so special-call on 09058094597

Congratulations ur awarded 500 of CD vouchers or 125gift guaranteed & Free entry 2 100 wkly draw txt MUSIC to 87066 TnCs www.Ldew.com1win150ppmx3age16

2. Dublin Institute of Technology (DIT)

Spam,'Sorry SMS missed you. We can deliver tickets on 10/04. Your ID and sig required. Pls confirm by TEXT with Y or N and new date and new address (NOT before 10/04)'

Spam,'FREEMSG: Our records indicate you may be entitled to 3750 pounds for the Accident you had. To claim for free reply YES to this msg.To opt out text STOP.'

Spam,'I\'m going to warn you once. If you dont stop txting& emailing my wife. I\'ll come round & sort u out.'

Spam,'Don\'t think u can be f***** clever wiv me I know what ur up to!'

Spam,'I\'m going to warn you once. If you dont stop txting& emailing my wife. I\'ll come round & sort u out'

Spam,'Let me jog your memory -Lizzie! Ring any bells? or is there more than one?'

Spam,'hello stranger.are you still as boring as ever. WB soon.xxx'

3. British English Corpora (BEC)

'PRIVATE! Your 2003 Account Statement for shows 800 un-redeemed S. I. M. points. Call 08715203652 Identifier Code: 42810 Expires 29/10/0'

'U have won a nokia 6230 plus a free digital camera. This is what u get when u win our FREE auction. To take part send NOKIA to 83383 now. POBOX114/14TCR/W1 16'

'Guess what! Somebody you know secretly fancies you! Wanna find out who it is? Give us a call on 09065394514 From Landline DATEBox1282EssexCM61XN 150p/min 18'

'Free video camera phones with Half Price line rental for 12 mths and 500 cross ntwk mins 100 txts. Call MobileUpd8 08001950382 or Call2OptOut/674&'

'Orange customer, you may now claim your FREE CAMERA PHONE upgrade for your loyalty. Call now on 0207 153 9996. Offer ends 14thMarch. T&C\'s apply. Opt-out availa'

'Free msg. Sorry, a service you ordered from 81303 could not be delivered as you do not have sufficient credit. Please top up to receive the service.'

'Want the latest Video handset? 750 anytime any network mins? Half price line rental? Reply or call 08000930705 for delivery tomorrow'

4. UCI Machine Learning (UCI)

Buy Space Invaders 4 a chance 2 win orig Arcade Game console. Press 0 for Games Arcade (std WAP charge) See o2.co.uk/games 4 Terms + settings. No purchase

Loan for any purpose -ú500 - -ú75,000. Homeowners + Tenants welcome. Have you been previously refused? We can still help. Call Free 0800 1956669 or text back 'help'

BIG BROTHER ALERT! The computer has selected u for 10k cash or #150 voucher. Call 09064018838. NTT PO Box CRO1327 18+ BT Landline Cost 150ppm mobiles vary

WIN: We have a winner! Mr. T. Foley won an iPod! More exciting prizes soon, so keep an eye on ur mobile or visit www.win-82050.co.uk

TodaysVoda numbers ending 1225 are selected to receive a -ú50award. If you have a match please call 08712300220 quoting claim code 3100 standard rates app

Hottest pics straight to your phone!! See me getting Wet and Wanting, just for you xx Text PICS to 89555 now! txt costs 150p textoperator g696ga 18 XxX

Hack Chat. Get backdoor entry into 121 chat rooms at a fraction of the cost. Reply NEO69 or call 09050280520, to subscribe 25p pm. DPS, Bcm box 8027 Ldn, wc1n3xx

5. SMSv.0.1

Urgent! call 09061749602 from Landline. Your complimentary 4* Tenerife Holiday or £10,000 cash await collection SAE T&Cs BOX 528 HP20 1YF 150ppm 18+,spam

+449071512431 URGENT! This is the 2nd attempt to contact U!U have WON £1250 CALL 09071512433 b4 050703 T&CsBCM4235WC1N3XX. callcost 150ppm mobilesvary. max£7. 50,spam

FREE for 1st week! No1 Nokia tone 4 ur mob every week just txt NOKIA to 8007 Get txting and tell ur mates www.getzed.co.uk POBox 36504 W45WQ norm150p/tone 16+,spam

Urgent! call 09066612661 from landline. Your complementary 4* Tenerife Holiday or £10,000 cash await collection SAE T&Cs PO Box 3 WA14 2PX 150ppm 18+ Sender: HolOffer,spam

WINNER!! As a valued network customer you have been selected to receivea £900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.,spam

07732584351 - Rodger Burns - MSG = We tried to call you re your reply to our sms for a free nokia mobile + free camcorder. Please call now 08000930705 for delivery tomorrow,spam

Appendix B

Detection phase

1. Keywords of ham and spam SMS messages for detection process

GROUP	LIST OF KEYWORDS
Spam	Free/ Terms and Conditions/Call/Dial/Tel/Landline/Click/Visit/Log onto/Direct Access/Gotto/Info/reply/charge/unsubscribe/subscribed/subscription/identifier code/discount code/claim code/valid/expires/offer end/update/urgent/private/ contact/helpline/ helpdesk/register/text/send/sent/claim/content/mobile content/ video content/mobile/phone/link/www/webpage/website/email/@/.com/cancel/ end/stop/rate/(/min)(/msg)/optout/offer/pound/redeem/guaranteed/customer service/
Ham	☺/ ☺ / =) / ☹ / ☹/ =(/ :D / :-D/ :-P/ :P/ ;-)/ :-xx/ Sunday/Monday/yuesday/Wednesday/Thursday/Friday/Saturday/today/yesterday/ tomorrow/day/year/weekend/sorry/thank/you/good/nice/fine/great/enjoy/fun/well done/hello/cheers/morning/afternoon/evening/night/noon/bye/nite/papa/mama/ mum/mom/mummy/mother/father/sister/brother/aunty/uncle/cousin/nephew/sir/ madam/Mr/Mrs/parent/family/guys/man/men/female/ladies/boy/girl/student/ person/people/he/him/his/she/her/hers/it/mine/my/they/them/their/myself/yourself /himself/herself/itself/ourselves/yourselves/themselves/lunch/dinner/breakfast/Ok /Okay/Okie/dokey/No/Nope/Yes/Sure/better/Alright/alrite/Right/Hey/ Hor/ Nah/ Lor/ Lar/ Lol/ Leh/ Oh/ Eh/ Wow/ Wah/Ahhh/ ah/ Aha/ aah/Argh/ Oops/ Oi/ Hi/ Erm/ Ish/ Er/ Yep/ Yeah/ Hiya/ Kinda/ Ow/ Ooh/ Ho/Hmm/ Yup/Wah/ shhhhh/ Mmmmmmm/ Hmmmmmm/ yummy/ Home/ Town/ Job/ Shop/ School/ Say/ Said/ Ask/ Asked/So/ Again/ Believe/ Yet/ Soon/ There/ here/ Maybe/ Immediately/ Unfortunately/ Fortunately/always/ already/ another/ anymore/ anyone/ anything/ anytime/ anywhere/whatever/wonder/everyone/somebody/ someday/ someone/something/ sometime/ somewhat/ somewhere/ together/ jogging/ shopping/ journey/ studying/ dancing/ joking/ driving/ training/ watching/sad/ happy/ excited/ angry/ lazy/ noise/ bored/ scary/ depresses/ lovely/ love/ quick/ quite/ slowly/ creepy/ smile/ scream/ shout/ silent/slow/ fast/quickly/ weak/easy/sweet dreams/ sweet/ sleep/ forgot/forget/ kind/beautiful/ ugly/ small/ big/ honey/ gently/ smell/ hot/ cool/ cold/ cute/ macho/ embarrass/…/..

2. Programming for detection process

i. Detection using Algorithm 1

a. Algorithm 1 Process 1

```
1  package detectionhamspam;
2  import java.io.*;
3  public class Algorithm1process1 {
4      public static void main (String args []){
5          // final long startTime = System.nanoTime();
6          try{
7          FileInputStream fstream = new FileInputStream ("UCISpam.txt");
8          PrintStream writer = new PrintStream ( new File ("UCITesting.txt"));
9
10         DataInputStream in = new DataInputStream (fstream);
11         BufferedReader br = new BufferedReader (new InputStreamReader (in));
12         String messages;
13
14         while ((messages = br.readLine()) !=null)
15         {
16             char [] charArray;
17             charArray = messages.toCharArray();
18
19             if (messages.contains("Free"))
20                     writer.println (messages + "\t -----> \tSPAM");
21             else if (messages.contains("free"))
22                     writer.println (messages + "\t -----> \tSPAM");
23             else if (messages.contains("FREE"))
24                     writer.println (messages + "\t -----> \tSPAM");
25
26             else if (messages.contains("TERMS AND CONDITIONS"))
27                     writer.println (messages + "\t -----> \tSPAM");
28             else if (messages.contains("TERMS & CONDITIONS"))
29                     writer.println (messages + "\t -----> \tSPAM");
30             else if (messages.contains("Terms and Conditions"))
31                     writer.println (messages + "\t -----> \tSPAM");
```

b. Algorithm 1 Process 2

```
package detectionhamspam;
import java.io.*;
public class Algorithm1process2 {
    public static void main (String args []){
      // final long startTime = System.nanoTime();
       try{
       FileInputStream fstream = new FileInputStream ("UCIHamclean_Algorithm1_output1_process1.txt");
       PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm1_output1_process2.txt"));

       DataInputStream in = new DataInputStream (fstream);
       BufferedReader br = new BufferedReader (new InputStreamReader (in));
       String messages;

       while ((messages = br.readLine()) !=null)
       {
           char [] charArray;
           charArray = messages.toCharArray();

           if (messages.contains("ToHam")&& messages.contains(":)"))
                    writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains(":-)"))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains("=)"))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains(":("))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains(":-("))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains("=("))
                   writer.println (messages + "\t -----> \tHAM");
```

ii. Detection using Algorithm 2

a. Algorithm 2 Process 1

```
1  package detectionhamspam;
2  import java.io.*;
3  public class Algorithm2process1 {
4      public static void main (String args []){
5          // final long startTime = System.nanoTime();
6          try{
7          FileInputStream fstream = new FileInputStream ("UCIHamclean.txt");
8          PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm2_output1_process1.txt"));
9
10         DataInputStream in = new DataInputStream (fstream);
11         BufferedReader br = new BufferedReader (new InputStreamReader (in));
12         String messages;
13
14         while ((messages = br.readLine()) !=null)
15         {
16             char [] charArray;
17             charArray = messages.toCharArray();
18
19             if ((messages.length () > 100 ) && messages.contains("free"))
20                     writer.println (messages + "\t ----> \tSPAM");
21             else if ((messages.length () > 100 ) && messages.contains("free"))
22                     writer.println (messages + "\t -----> \tSPAM");
23             else if ( (messages.length () > 100 ) && messages.contains("FREE"))
24                     writer.println (messages + "\t ----> \tSPAM");
25
26             else if ( (messages.length () > 100 ) && messages.contains("TERMS AND CONDITIONS"))
27                     writer.println (messages + "\t -----> \tSPAM");
28             else if ((messages.length () > 100 ) && messages.contains("TERMS & CONDITIONS"))
29                     writer.println (messages + "\t -----> \tSPAM");
30             else if ((messages.length () > 100 ) && messages.contains("Terms and Conditions"))
31                     writer.println (messages + "\t -----> \tSPAM");
```

b. Algorithm 2 Process 2

```
1  package detectionhamspam;
2  import java.io.*;
3  public class Algorithm2process2 {
4
5      public static void main (String args []){
6      // final long startTime = System.nanoTime();
7       try{
8       FileInputStream fstream = new FileInputStream ("UCIHamclean_Algorithm2_output1_process1.txt");
9       PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm2_output1_process2.txt"));
10
11      DataInputStream in = new DataInputStream (fstream);
12      BufferedReader br = new BufferedReader (new InputStreamReader (in));
13      String messages;
14
15      while ((messages = br.readLine()) !=null)
16      {
17          char [] charArray;
18          charArray = messages.toCharArray();
19
20          if (messages.contains("ToHam")&& messages.contains(":)"))
21                  writer.println (messages + "\t -----> \tHAM");
22          else if (messages.contains("ToHam") && messages.contains(":-)"))
23                  writer.println (messages + "\t -----> \tHAM");
24          else if (messages.contains("ToHam") && messages.contains("=)"))
25                  writer.println (messages + "\t -----> \tHAM");
26          else if (messages.contains("ToHam") && messages.contains(":("))
27                  writer.println (messages + "\t -----> \tHAM");
28          else if (messages.contains("ToHam") && messages.contains(":-("))
29                  writer.println (messages + "\t -----> \tHAM");
30          else if (messages.contains("ToHam") && messages.contains("=("))
31                  writer.println (messages + "\t -----> \tHAM");
```

iii. Detection using Algorithm 3

a. Algorithm 3 Process 1

```
package detectionhamspam;
import java.io.*;
public class Algorithm3process1 {
    public static void main (String args []){
       // final long startTime = System.nanoTime();
        try{
        FileInputStream fstream = new FileInputStream ("UCIHamclean.txt");
        PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm3_output1_process1.txt"));

        DataInputStream in = new DataInputStream (fstream);
        BufferedReader br = new BufferedReader (new InputStreamReader (in));
        String messages;

        while ((messages = br.readLine()) !=null)
        {
            char [] charArray;
            charArray = messages.toCharArray();

            if ((messages.matches(".*\\d+.*")) && messages.contains("Free"))
                    writer.println (messages + "\t -----> \tSPAM");
            else if ((messages.matches(".*\\d+.*")) && messages.contains("free"))
                    writer.println (messages + "\t -----> \tSPAM");
            else if ((messages.matches(".*\\d+.*")) && messages.contains("FREE"))
                    writer.println (messages + "\t -----> \tSPAM");

            else if ((messages.matches(".*\\d+.*")) && messages.contains("TERMS AND CONDITIONS"))
                    writer.println (messages + "\t -----> \tSPAM");
            else if ((messages.matches(".*\\d+.*")) && messages.contains("TERMS & CONDITIONS"))
                    writer.println (messages + "\t -----> \tSPAM");
            else if ((messages.matches(".*\\d+.*")) && messages.contains("Terms and Conditions"))
                    writer.println (messages + "\t -----> \tSPAM");
```

b. Algorithm 3 Process 2

```
package detectionhamspam;
import java.io.*;
public class Algorithm3process2 {
    public static void main (String args []){
       // final long startTime = System.nanoTime();
        try{
        FileInputStream fstream = new FileInputStream ("UCIHamclean_Algorithm3_output1_process1.txt");
        PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm3_output1_process2.txt"));

        DataInputStream in = new DataInputStream (fstream);
        BufferedReader br = new BufferedReader (new InputStreamReader (in));
        String messages;

        while ((messages = br.readLine()) !=null)
        {
            char [] charArray;
            charArray = messages.toCharArray();

            if (messages.contains("ToHam") && messages.contains(":)"))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam") && messages.contains(":-)"))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam") && messages.contains("=)"))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam") && messages.contains(":("))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam") && messages.contains(":-("))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam") && messages.contains("=("))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam") && messages.contains(":D"))
```

iv. Detection using Algorithm 4

a. Algorithm 4 Process 1

```
1  package detectionhamspam;
2  import java.io.*;
3  public class Algorithm4process1 {
4      public static void main (String args []){
5        // final long startTime = System.nanoTime();
6         try{
7         FileInputStream fstream = new FileInputStream ("UCIHamclean.txt");
8         PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm4_output(process1.txt"));
9
10        DataInputStream in = new DataInputStream (fstream);
11        BufferedReader br = new BufferedReader (new InputStreamReader (in));
12        String messages;
13
14        while ((messages = br.readLine()) !=null)
15        {
16            char [] charArray;
17            charArray = messages.toCharArray();
18            if (messages.length () > 100 )
19                    writer.println (messages + "\t ----> \tTrue");
20
21              else
22
23                writer.println (messages + "\t ----> \tHAM");
24
25        }
26        in.close();
27        writer.close();
28        }
29        catch (Exception e)
30        {
31            System.out.println ("Error");
```

b. Algorithm 4 Process 2

```
package detectionhamspam;
import java.io.*;
public class Algorithm4process2 {
    public static void main (String args []){
      // final long startTime = System.nanoTime();
       try{
       FileInputStream fstream = new FileInputStream ("UCIHamclean_Algorithm4_output1_process1.txt");
       PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm4_output1_process2.txt"));

       DataInputStream in = new DataInputStream (fstream);
       BufferedReader br = new BufferedReader (new InputStreamReader (in));
       String messages;

       while ((messages = br.readLine()) !=null)
       {
           char [] charArray;
           charArray = messages.toCharArray();

           if ((messages.matches(".*\\d+.*")) && messages.contains("Free")&& messages.contains("ToSpam"))
                   writer.println (messages + "\t -----> \tSPAM");
           else if ((messages.matches(".*\\d+.*")) && messages.contains("free")&& messages.contains("ToSpam"))
                  writer.println (messages + "\t -----> \tSPAM");
           else if ( (messages.matches(".*\\d+.*")) && messages.contains("FREE")&& messages.contains("ToSpam"))
                  writer.println (messages + "\t -----> \tSPAM");

           else if ( (messages.matches(".*\\d+.*")) && messages.contains("TERMS AND CONDITIONS")&& messages.contains("ToSpam"))
                  writer.println (messages + "\t -----> \tSPAM");
           else if ((messages.matches(".*\\d+.*")) && messages.contains("TERMS & CONDITIONS")&& messages.contains("ToSpam"))
                  writer.println (messages + "\t -----> \tSPAM");
           else if ((messages.matches(".*\\d+.*")) && messages.contains("Terms and Conditions")&& messages.contains("ToSpam"))
                  writer.println (messages + "\t -----> \tSPAM");
```

c. Algorithm 4 Process 3

```
package detectionhamspam;
import java.io.*;
public class Algorithm4process3 {
    public static void main (String args []){
      // final long startTime = System.nanoTime();
       try{
       FileInputStream fstream = new FileInputStream ("UCIHamclean_Algorithm4_output1_process2.txt");
       PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm4_output1_process3.txt"));

       DataInputStream in = new DataInputStream (fstream);
       BufferedReader br = new BufferedReader (new InputStreamReader (in));
       String messages;

       while ((messages = br.readLine()) !=null)
       {
           char [] charArray;
           charArray = messages.toCharArray();

           if (messages.contains("ToHam")&& messages.contains(":)"))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains(";-"))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains("=)"))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains(":("))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains(":-("))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains("=("))
                   writer.println (messages + "\t -----> \tHAM");
           else if (messages.contains("ToHam") && messages.contains(":D"))
```

v. Detection using Algorithm 5

a. Algorithm 5 Process 1

```
1   package detectionhamspam;
2   import java.io.*;
3   public class Algorithm5process1 {
4       public static void main (String args []){
5          // final long startTime = System.nanoTime();
6           try{
7           FileInputStream fstream = new FileInputStream ("UCIHamclean.txt");
8           PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm5_output1_process1.txt"));
9
10          DataInputStream in = new DataInputStream (fstream);
11          BufferedReader br = new BufferedReader (new InputStreamReader (in));
12          String messages;
13
14          while ((messages = br.readLine()) !=null)
15          {
16              char [] charArray;
17              charArray = messages.toCharArray();
18              if (messages.length () > 100 )
19                      writer.println (messages + "\t -----> \tTxtHam");
20
21               else
22
23                 writer.println (messages + "\t -----> \tNoHam1");
24
25          }
26          in.close();
27          writer.close();
28          }
29          catch (Exception e)
30          {
31              System.out.println ("Error");
```

b. Algorithm 5 Process 2

```
package detectionhamspam;
import java.io.*;
public class Algorithm5process2 {
    public static void main (String args []){
      // final long startTime = System.nanoTime();
       try{
       FileInputStream fstream = new FileInputStream ("UCIHamclean_Algorithm5_output1_process1.txt");
       PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm5_output1_process2.txt"));

       DataInputStream in = new DataInputStream (fstream);
       BufferedReader br = new BufferedReader (new InputStreamReader (in));
       String messages;

       while ((messages = br.readLine()) !=null)
       {
          char [] charArray;
          charArray = messages.toCharArray();

          if ((messages.matches(".*\\d+.*")) && messages.contains("Free")&& messages.contains("ToSpam"))
                writer.println (messages + "\t -----> \tSPAM");
          else if ((messages.matches(".*\\d+.*")) && messages.contains("free")&& messages.contains("ToSpam"))
                writer.println (messages + "\t -----> \tSPAM");
          else if ( (messages.matches(".*\\d+.*")) && messages.contains("FREE")&& messages.contains("ToSpam"))
                writer.println (messages + "\t -----> \tSPAM");

          else if ( (messages.matches(".*\\d+.*")) && messages.contains("TERMS AND CONDITIONS")&& messages.contains("ToSpam"))
                writer.println (messages + "\t -----> \tSPAM");
          else if ((messages.matches(".*\\d+.*")) && messages.contains("TERMS & CONDITIONS")&& messages.contains("ToSpam"))
                writer.println (messages + "\t -----> \tSPAM");
          else if ((messages.matches(".*\\d+.*")) && messages.contains("Terms and Conditions")&& messages.contains("ToSpam"))
                writer.println (messages + "\t -----> \tSPAM");
```

c. Algorithm 5 Process 3

```
package detectionhamspam;
import java.io.*;
public class Algorithm5process3 {
    public static void main (String args []){
      // final long startTime = System.nanoTime();
       try{
       FileInputStream fstream = new FileInputStream ("UCIHamclean_Algorithm5_output1_process2.txt");
       PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm5_output1_process3.txt"));

       DataInputStream in = new DataInputStream (fstream);
       BufferedReader br = new BufferedReader (new InputStreamReader (in));
       String messages;

       while ((messages = br.readLine()) !=null)
       {
           char [] charArray;
           charArray = messages.toCharArray();

            if ( messages.contains("WWW")&& messages.contains("ToHam1"))
                  writer.println (messages + "\t -----> \tSPAM");
           else if (messages.contains("www")&& messages.contains("ToHam1"))
                  writer.println (messages + "\t -----> \tSPAM");
           else if ( messages.contains("http")&& messages.contains("ToHam1"))
                  writer.println (messages + "\t -----> \tSPAM");
           else if ( messages.contains("HTTP")&& messages.contains("ToHam1"))
                  writer.println (messages + "\t -----> \tSPAM");

           else if ( messages.contains("@")&& messages.contains("ToHam1"))
                  writer.println (messages + "\t -----> \tSPAM");
           else if ( messages.contains(".com")&& messages.contains("ToHam1"))
                  writer.println (messages + "\t -----> \tSPAM");
```

d. Algorithm 5 Process 4

```
package detectionhamspam;
import java.io.*;
public class Algorithm5process4 {
    public static void main (String args []){
       // final long startTime = System.nanoTime();
        try{
        FileInputStream fstream = new FileInputStream ("UCIHamclean_Algorithm5_output1_process3.txt");
        PrintStream writer = new PrintStream ( new File ("UCIHamclean_Algorithm5_output1_process4.txt"));

        DataInputStream in = new DataInputStream (fstream);
        BufferedReader br = new BufferedReader (new InputStreamReader (in));
        String messages;

        while ((messages = br.readLine()) !=null)
        {
            char [] charArray;
            charArray = messages.toCharArray();

            if (messages.contains("ToHam2")&& messages.contains(":)"))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam2") && messages.contains(":-)"))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam2") && messages.contains("=)"))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam2") && messages.contains(":("))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam2") && messages.contains(":-("))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam2") && messages.contains("=("))
                    writer.println (messages + "\t -----> \tHAM");
            else if (messages.contains("ToHam2") && messages.contains(":D"))
```

3. Example of ham and spam messages

CLASS	EXAMPLE
Spam	REMINDER FROM O2: To get 2.50 pounds free call credit and details of great offers pls reply 2 this text with your valid name, house no and postcode
	This is the 2nd time we have tried 2 contact u. U have won the -£750 Pound prize. 2 claim is easy, call 087187272008 NOW1! Only 10p per minute. BT-national-rate.
	December only! Had your mobile 11mths+? You are entitled to update to the latest colour camera mobile for Free! Call The Mobile Update Co FREE on 08002986906
	Do you want a New Nokia 3510i colour phone DeliveredTomorrow? With 300 free minutes to any mobile + 100 free texts + Free Camcorder reply or call 08000930705.
	PRIVATE! Your 2004 Account Statement for 078498****7 shows 786 unredeemed Bonus Points. To claim call 08719180219 Identifier Code: 45239 Expires 06.05.05
Ham	Ok not a problem will get them a taxi. C ing tomorrow and tuesday. On tuesday think we r all going to the cinema.
	sorry, no, have got few things to do. may be in pub later.
	Our ride equally uneventful - not too many of those pesky cyclists around at that time of night ;).
	Ha ha - had popped down to the loo when you hello-ed me. Hello!
	By the way, i've put a skip right outside the front of the house so you can see which house it is. Just pull up before it.

Appendix C

Classification phase

1. Keywords for clustering phase in each cluster

METHOD	CATEGORY	SUB-CATEGORY	
Keywords	Divide the messages into 16 categories		
	Financial	Cash-balance/Cash-in/Account Statement/Loan/ Credit card/ Debts/ Finance	
	Offer	Offer/Half price (1/2 price)/ Activate	
	Prize	Reward/Selected/Prize/Award/Guaranteed/Holiday/Congratulation/ Lucky	
	Service	Delivery/Update	
	Games/ Competition	To win/Correct answer/ Quiz/Answer/Correct/Incorrect/Competition/ Question/Free entry/Quiz Weekly/ Weekly comp/Take part	
	Ringtone	Ringtone/Tone/Poly	
	Chat	Chat/ Friend/ Gender/Name/Age/Adult/Married/Secret admire/Fancies/ Female	
	Date	Date/Blind /Area	
	Sex	Sex/Dogging/Horny/Naked/Hot live Fantasies/Fantasy/Gay/Flirt/Filthy/ Cook/Porn/Gigolo/Passion/Sexual/Alone/Wet/Smashed/Nasty/Fancy/ Knickers	
	VoiceMail	Voicemail/Urgent message/ You have 1 new message	
	MailMessages	Mailbox messaging/Mailbox	
	Claim	Compensation/Accident	
	Entertainment	Video	
	Miscellaneous	Messages that cannot be categorized will be clustered again with sub-category	
		Financial	Bill/Insurance/Paid/Debt/Card/Point/Money
		Offer	2/2 register/Discount code/Contract/Wallpaper
		Prize	Voucer/Topped/Won/Congrat/Price
		Service	Service/Help/Click/Direct access/Order/Upgrade/ Accomondation
		Games/ Competition	Game/Collect/Play/Good luck/A-/Participate/Win/ Bid
		Chat	Mind/Heart/Care/Darling/Talk/Number/Touch/Name /Landline/Dirty/Fanies
		Date	Perfect partner/match/girls
		Sex	Home/Lady/Naughty/Bed/Brunette/Hot/Penetrate
		Voicemail	Alert
		MailMessages	You have a free message/You have a new message
		Job	Job/cv/Company
		Claim	Entitled
		Advertisement	Sale/Goal/News/Dream/Find out/Amazing/Donate/ Garden
		Entertainment	Xmas/Special/Romantic/Picture/Music/Sing/Fun/ Content
		Other	Messages that cannot be clustered

2. Clustering using HICNA

```
package clusteringspam;
import java.io.*;
public class clusteringSpam5 {
    public static void main (String args []){
       // final long startTime = System.nanoTime();
        try{
        FileInputStream fstream = new FileInputStream ("BritishEnglishRawSMS.txt");
        PrintStream writer = new PrintStream ( new File ("BritishEnglishRawSMS_RawSMS1_result.txt"));

        DataInputStream in = new DataInputStream (fstream);
        BufferedReader br = new BufferedReader (new InputStreamReader (in));
        String messages;

        while ((messages = br.readLine()) !=null)
        {
            char [] charArray;
            charArray = messages.toCharArray();

          if (messages.contains("CASH-BALANCE"))
                writer.println (messages + "\t -----> \tFINANCIAL");
           else if (messages.contains("Cash-Balance"))
                writer.println (messages + "\t -----> \tFINANCIAL");
            else if (messages.contains("Cash-balance"))
                writer.println (messages + "\t -----> \tFINANCIAL");
            else if (messages.contains("cash-balance"))
                writer.println (messages + "\t -----> \tFINANCIAL");
           else if (messages.contains("CASH-IN"))
                writer.println (messages + "\t -----> \tFINANCIAL");
```

3. Example of spam messages in each cluster

CLUSTER	EXAMPLE OF MESSAGES
Financial	PRIVATE! Your 2003 Account Statement for shows 800 un-redeemed S. I. M. points. Call 08718738002 Identifier Code: 48922 Expires 21/11/04
	Ur cash-balance is currently 500 pounds - to maximize ur cash-in now send COLLECT to *83600 only 150p/msg. CC: 08718720201* PO BOX 114/14 TCR/W1
	Your bill payment request of TK 700.00 (Transaction id BD8051111360356) has been successful. Thank you. Please call 1211 to know your latest account info.
Offer	Update_Now - Xmas Offer! Latest Motorola, SonyEricsson & Nokia & FREE Bluetooth! Double Mins & 1000 Txt on Orange. Call MobileUpd8 on 08000839402 or call2optout/F4Q=
	REMINDER FROM O2: To get 2.50 pounds free call credit and details of great offers pls reply 2 this text with your valid name, house no and postcode
	We tried to contact you re our offer of New Video Phone 750 anytime any network mins HALF PRICE Rental camcorder call 08000930705 or reply for delivery Wed
Prize	WINNER!! As a valued network customer you have been selected to receivea £900 prize reward! To claim call 09061701461. Claim code KL341. Valid 12 hours only.
	URGENT! Your mobile number ********** WON a £2000 Bonus Caller prize on 10/06/03! This is the 2nd attempt to reach you! Call 09066368753 ASAP! Box 97N7QP, 150ppm
	Todays Voda numbers ending 7548 are selected to receive a $350 award. If you have a match please call 08712300220 quoting claim code 4041 standard rates app
Service	!!URGENT!! Large parcel awaiting delivery Please call now on +43820899510 for delivery tomorrow Regards International parcel deliveries
	Customer service annonc**e**ment. You have a New Years delivery waiting for you. Please call 07046759868 now to arrange delivery
	You have an important customer service announcement. Call FREEPHONE 0800 542 0825 now!
Games/ Competition	Free entry to the gr8prizes wkly comp 4 a chance to win the latest Nokia 8800, PSP or £250 cash every wk. TXT GREAT to 80878 www.gr8prizes.com 08715705022
	Sunshine Quiz Wkly Q! Win a top Sony DVD player if u know which country the Algarve is in? Txt ans to 82277. £1.50 SP:Tyrone
	Ur balance is now £500. Ur next question is: Who sang 'Uptown Girl' in the 80's ? 2 answer txt ur ANSWER to 83600. Good luck!
Ringtone	Thanks for your subscription to Ringtone UK your mobile will be charged £5/month Please confirm by replying YES or NO. If you reply NO you will not be charged
	Download as many ringtones as u like no restrictions, 1000s 2 choose. U can even send 2 yr buddys. Txt Sir to 80082 £3
	Get ur 1st RINGTONE FREE NOW! Reply to this msg with TONE. Gr8 TOP 20 tones to your phone every week just £1.50 per wk 2 opt out send STOP 08452810071 16
Chat	Dear U've been invited to XCHAT. This is our final attempt to contact u! Txt CHAT to 86688 150p/MsgrcvdHG/Suite342/2Lands/Row/W1J6HL LDN 18 yrs
	U have a secret admirer who is looking 2 make contact with U-find out who they R*reveal who thinks UR so special-call on 09058094565
	I don't know u and u don't know me. Send CHAT to 86688 now and let's find each other! Only 150p/Msg rcvd. HG/Suite342/2Lands/Row/W1J6HL LDN. 18 years or over.

Date	Someone U know has asked our dating service 2 contact you! Cant Guess who? CALL 09058091854 NOW all will be revealed. PO BOX385 M6 6WU
	We have new local dates in your area - Lots of new people registered in YOUR AREA. Reply DATE to start now! 18 only www.flirtparty.us REPLYS150
	Promotion Number: someone you know - U R being contacted by our dating service by someone you know! To find out who it is, call from a land line 09058099543. PoBox82W2QF150P
Sex	Im so Horny!!! Havnt had it in ages! Tell me what ud do to me right now! I need to be satisfied. Plz txt bac quick XXS
	sexy sexy cum and text me im wet and warm and ready for some porn! u up for some fun? THIS MSG IS FREE RECD MSGS 150P INC VAT 2 CANCEL TEXT STOP
	This message is free. Welcome to the new & improved Sex & Dogging club! To unsubscribe from this service reply STOP. msgs@150p 18+only
VoiceMail	Please CALL 08712402779 immediately as there is an urgent message waiting for you
	Promotion Number: 8712104073 - You have 1 new voicemail message. Please call 08712104073 to listen to it
	An Answer Phone message from <PHONENUMBER> was left for you at 10.12 on 26/03. Please call 123 to listen to it before it is automatically deleted.
MailMessage	<Forwarded from 21870000>Hi - this is your Mailbox Messaging SMS alert. You have 40 matches. Please call back on 09056242159 to retrieve your messages and matches cc100p/min
Claim	FREEMSG: Our records indicate you may be entitled to 3750 pounds for the Accident you had. To claim for free reply with YES to this msg. To opt out text STOP
Entertainment	Text82228>> Get more ringtones, logos and games from www.txt82228.com. Questions: info@txt82228.co.uk
	Get a FREE mobile video player FREE movie. T6 collect text GO to 89105. Its free! Extra films can be ordered t's and c's apply. 18 yrs only
	Twinks, bears, scallies, skins and jocks are calling now. Don't miss the weekend's fun. Call 08712466669 at 10p/min. 2 stop texts call 08712460324(nat rate)
Advertisement	sports fans - get the latest sports news str*,2 ur mobile 1 wk FREE PLUS a FREE TONE Txt SPORT ON to 8007 www.getzed.co.uk 0870141701216+ norm 4txt/120p
	For sale - arsenal dartboard. Good condition but no doubles or trebles!
	18 days to Euro2004 kickoff! U will be kept informed of all the latest news and results daily. Unsubscribe send GET EURO STOP to 83222.,spam
Job	Good morning i have found your cv on the internet and was wondering if you are still looking for a job. Please let me know asap. Thanks
	Hi Nicholas,this is Tameside College. Do u want 2 work as a Chef, earn a wage and study only 1 day a week?If yes call 01619086705 and ask for Maxine.
	pls call Richard Hadnum at Evolution on 0192583110 about an outstanding PERM web PHP (drupal) job in Macclesfield ref# 165876 or reply to text
Other	Do you like c*ks if so u cn earn ?200 2day cash
	Old man driving on the highway, phone rang. Answering, "Herman, there's a car going the wrong way on M60. Be careful!"-"It's not just one car, it'shundreds!"
	Did you hear about the new "Divorce Barbie"? It comes with all of Ken's stuff!

Appendix D

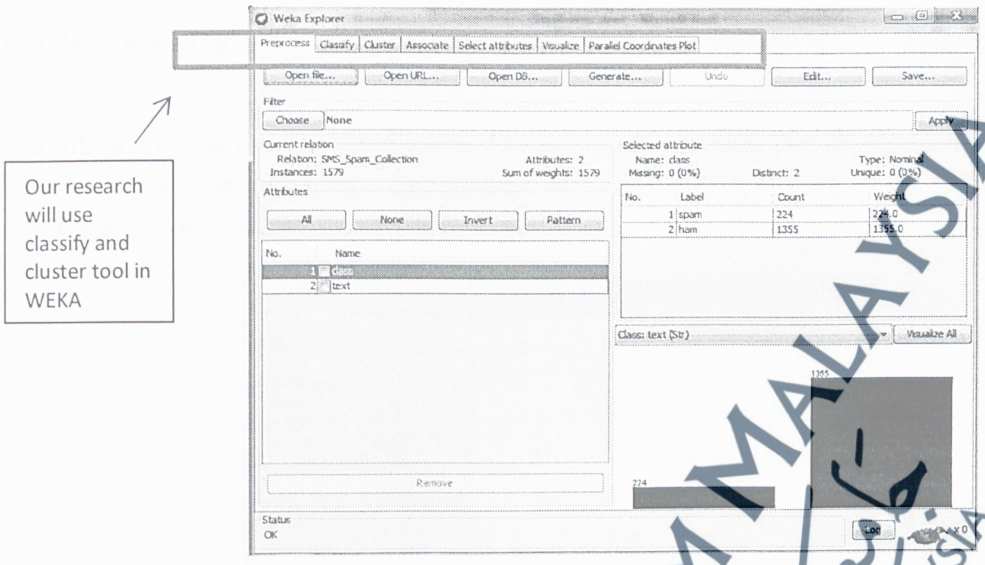
Detection and Clustering using WEKA

WEKA is a collection of machine learning algorithms for data mining tasks (Witten and Frank, 2005). The algorithms can either be applied directly to a dataset or called from your own Java code. WEKA stands for Waikato Environment for Knowledge Learning and it was developed by the University of Waikato, New Zealand. WEKA contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. The supported data formats in WEKA are ARFF, CSV, C4.5 and binary.

Front view in WEKA tool

There are four interfaces available in WEKA:-

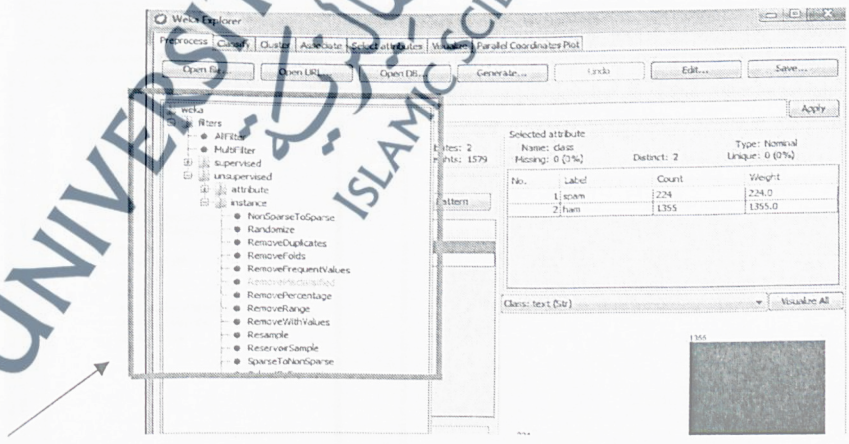
- Explorer – Environment for exploring data with WEKA. It gives access to all the facilities using menu selection and form filling.
- Experimenter – testing and evaluating machine learning algorithms. It can be used to get the answer for question: Which methods and parameter values work best for the given problem?
- Knowledge Flow – Same function as explorer. Supports incremental learning. It allows designing configurations for streamed data processing. Incremental algorithms can be used to process very large datasets
- Simple Command line – A simple interface for typing commands



Tools available in explorer interface

a. Preprocess in WEKA

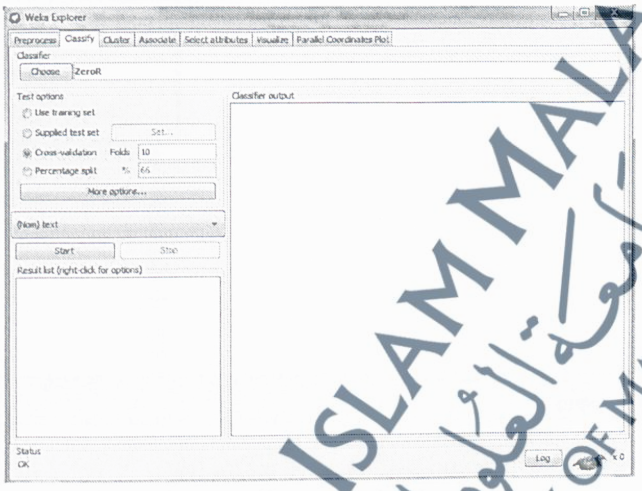
Preprocessing is the first step that we should do when we insert dataset in WEKA. We can remove any unused value or attributes in dataset or we want to change the character of dataset from string to binary. Besides that, it also can help WEKA run the dataset. For example, in classification tool, there are some algorithms available such as decision tress and support vector machine. Sometimes these algorithms cannot support the dataset that we used so that WEKA cannot run the data. By using preprocessing, we can change the character or value of dataset by applying filtering in preprocessing.



Filters in Preprocessing section

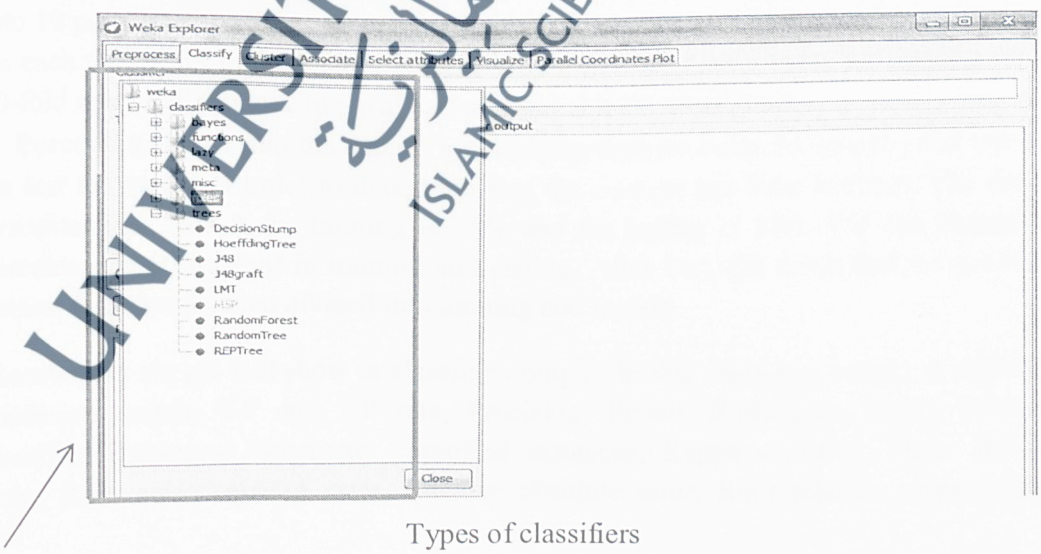
b. Classification in WEKA

Classification is supervised learning and require training data. The training set containing data that have been previously categorized. So classification is used to predict the target that must be categorical.



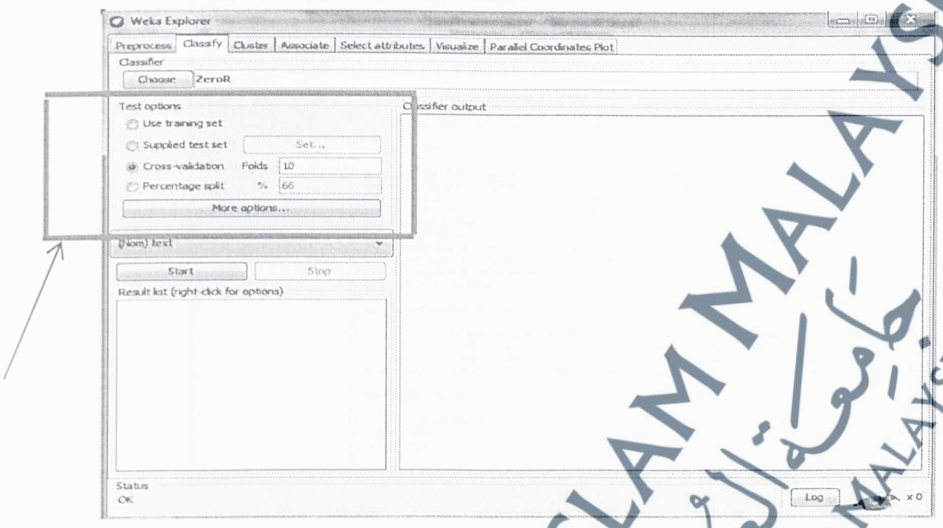
Interface in classification

Some classifier algorithms available in WEKA are bayes, functions, lazy, meta, misc, rules and trees. All of these classifiers are used to classify dataset. We cannot run WEKA without choosing classifiers. There are four classifiers that most popular and always be used by researchers; Decision Tree (Trees), Naïve Bayes (Bayes), Support Vector Machine (Functions) and Nearest Neighbour (Lazy).



Types of classifiers

Another role that we must do before classifying process is by choosing test option. The function of test options is to identify how dataset is tested. Four types of test option are Use training set, Supplied test set, Cross-validation and Percentage split.

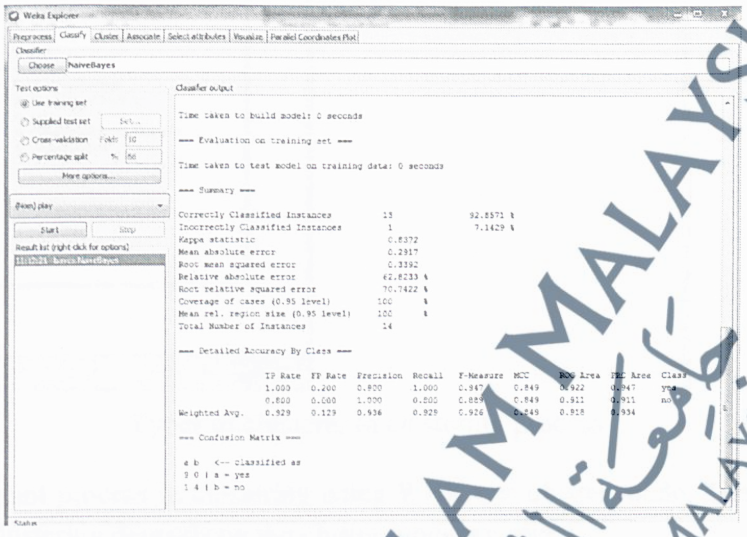


Interface for test options

- **Use training set**: Evaluates the classifiers on how well it predicts the class of the instances it was trained on. Use training dataset
- **Supplied test set**: to test again the prediction of the instances from training set to get more accuracy. Use training dataset
- **Cross-validation**: Process of repeated holdout that can reduce the variance of the estimate. The default of cross-validation in WEKA is 10, means that the dataset is divided into 10 pieces, 9 pieces use for training and the last piece for testing. Divide the dataset into 10 parts (these are called "folds"), hold out each part in turn, and average the results. So each data point in the dataset is used once for testing and 9 times for training. That's 10-fold cross-validation. Cross-validation is suitable for small data
- **Percentage split**: Run the dataset into training data (to build the model) and test data (to test the model in order to determine that the concept has been learned). The default percentage in WEKA for training is 66% and for testing is 34%. We can change the percentage that we want in training and testing. After that, the result that we get is the dataset that already been divided into training and testing.

Results that we get will show in classifiers output. In this classifiers output, it will show confusion matrix, TP rate, FP rate, Precision, Recall, F-Measure, MCC, Correctly classified instances, Incorrectly classified instances, Kappa statistics, Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error,

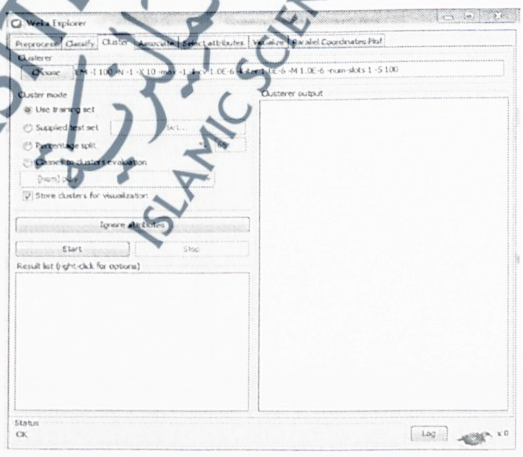
Coverage of cases, Mean rel. region size, Total number of instances, ROC area, PRC area and Time taken to build model.



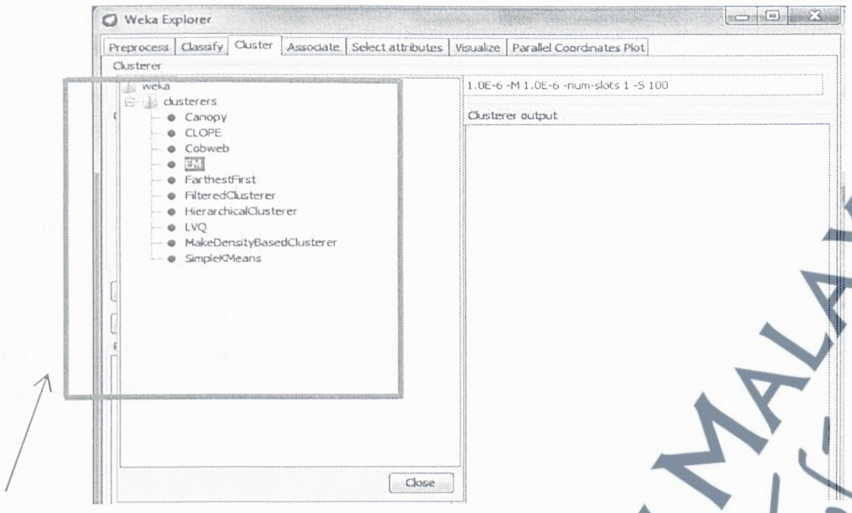
Interface for outputs in classification process

c. Clustering in WEKA

Clustering is unsupervised learning and it is a process of making group data into similar classes. Cluster is a group of objects that belong to the same class. Clusterer algorithms that available in clustering process are Canopy, CLOPE, Cobweb, EM, FarthestFirst, FilteredClusterer, HierarchicalClusterer, LVQ, MakeDensityBasedCluster and SimpleKMeans.

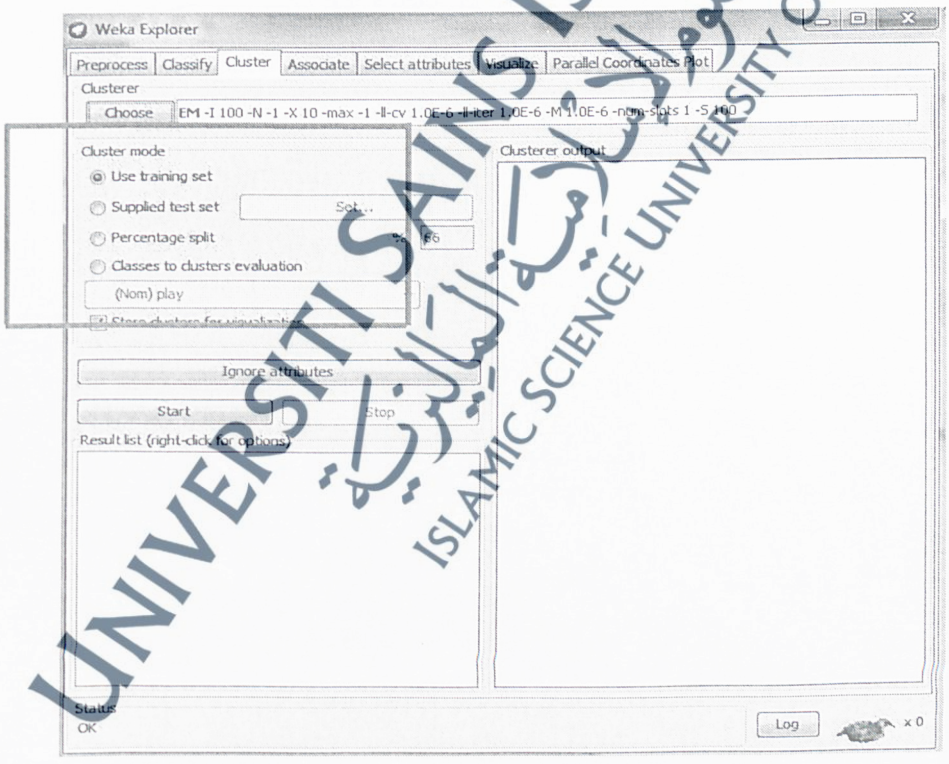


Interface in clustering



Types of clusterer in clustering process

Another important process in clustering using WEKA is cluster mode. The way WEKA evaluates the clustering depends on the cluster mode we select.



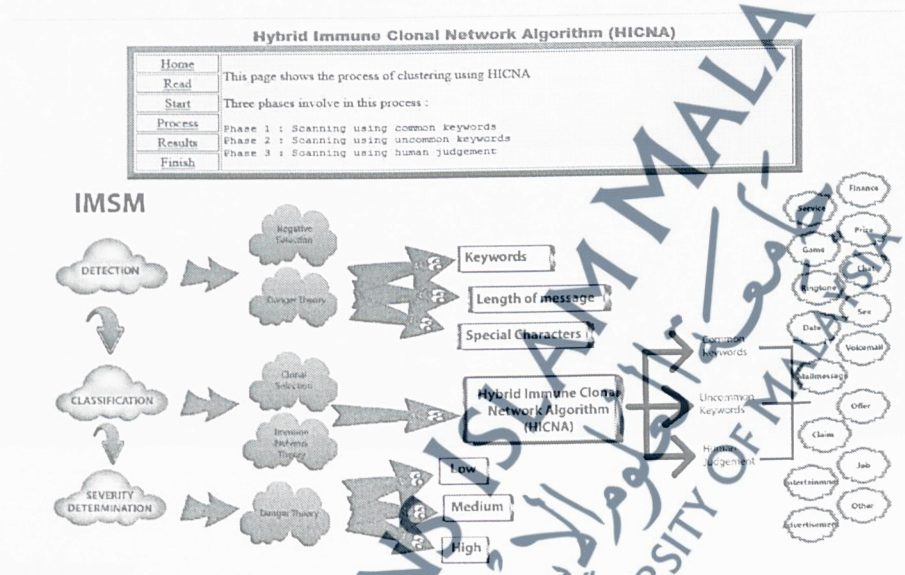
Cluster mode

- Use training set: After generating the clustering, WEKA classifies the training instances into clusters according to the cluster representation and computes the percentage of instances falling in each cluster. For example, when we used k-means for clustering with number of cluster is 2, result shows 43% (6 instances) in cluster 0 and 57% (8 instances) in cluster 1.
- Supplied test set: same function in classification
- Percentage split: same function in classification
- Classes to clusters evaluation: we select one attribute in dataset (for example, 'status'). Then we will compare how well the chosen clusters match up with a pre-assigned class ('status') in the data

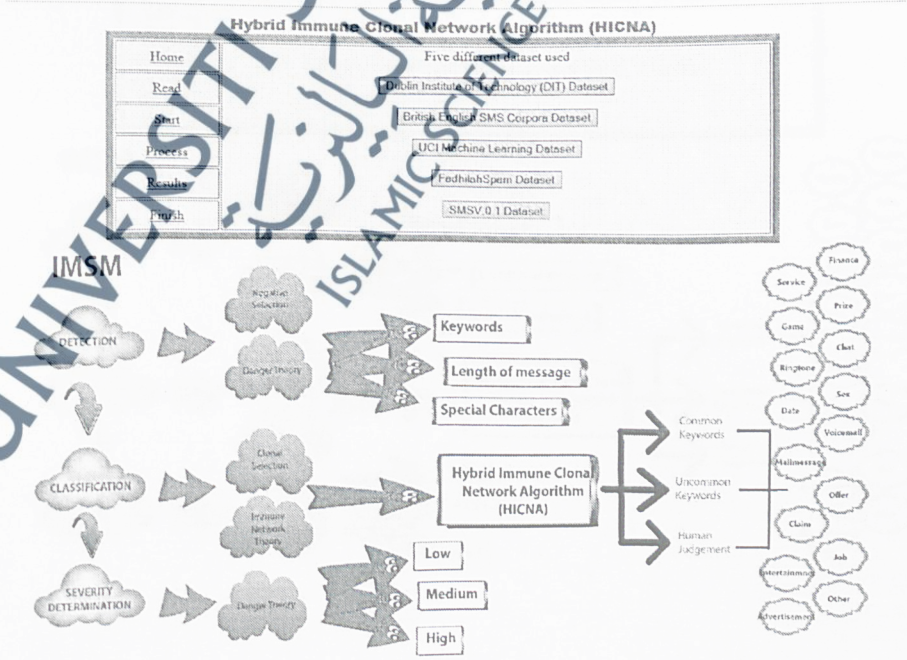
Appendix E

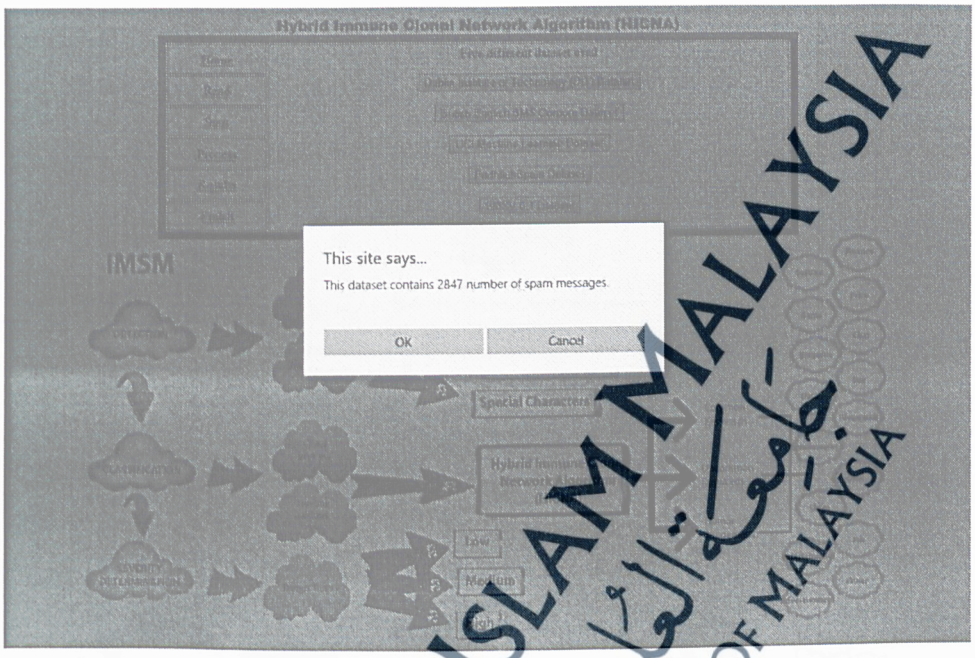
Interface for clustering application

a. Homepage

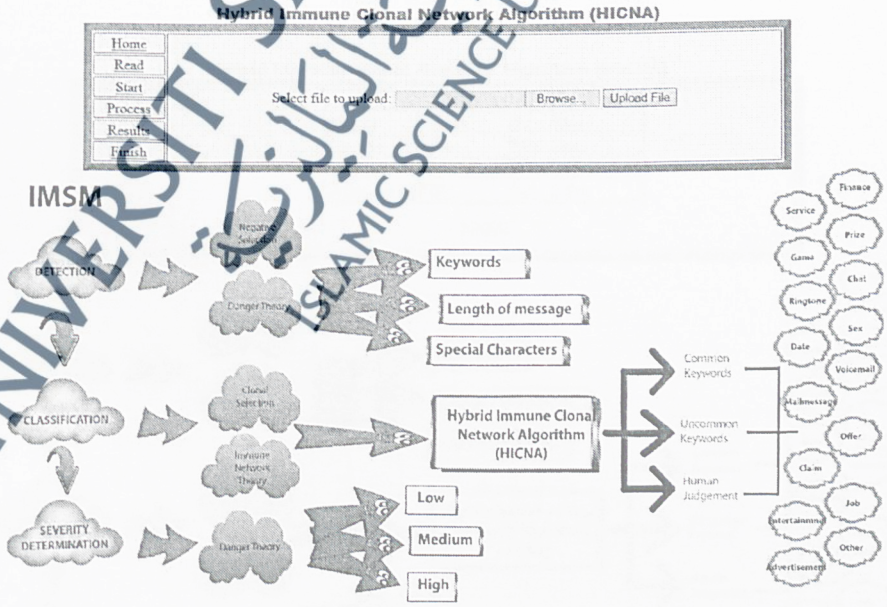


b. Read dataset

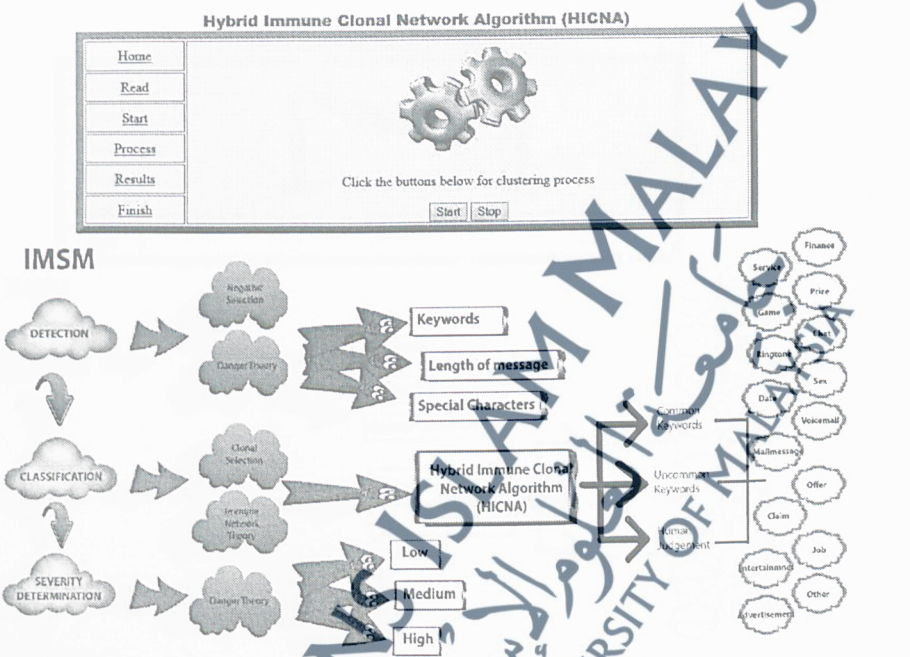




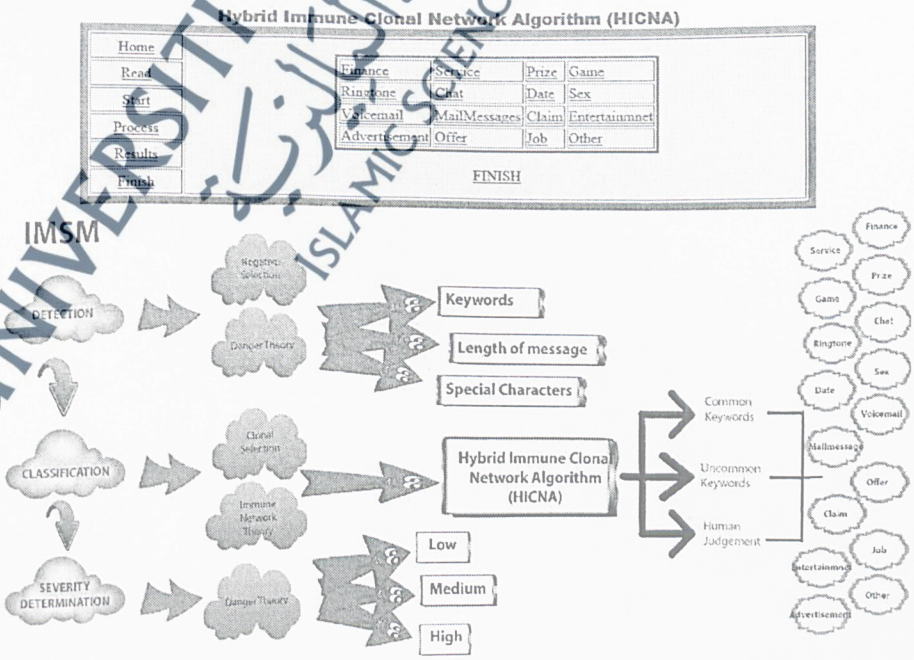
c. Upload dataset



d. Process of clustering



e. Results



f. Finish the process