

APPENDIX



Doc vs Internet + Library

97.05% Originality	2.95% Similarity	66 Sources
--------------------	------------------	------------

Web sources: 62 sources found

1. https://www.cambridge.org/core/journals/public-health-nutrition/article/risk-for-chronic-kidney-diseas...	1.16%
2. https://experts.umn.edu/en/publications/the-influence-of-gender-and-other-patient-characteristics-o...	0.75%
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4815064	0.75%
4. https://link.springer.com/content/pdf/10.1186%2Fs12875-016-0440-0.pdf	0.75%
5. http://www.hmonglibrary.org/hmong-studies-virtual-library---dissertations-and-theses.html	0.26%
6. https://www.outreach.usda.gov/sdfr/directory.htm	0.26%
7. http://www.fke.utm.my/eess2016/list-of-abstracts.php	0.26%
8. https://id.123dok.com/document/lzgjev8z-pengaruh-kompensasi-dan-lingkungan-kerja-pada-kinerja...	0.26%
9. http://www.hmonglibrary.org/hmong-studies-virtual-library.html	0.26%
10. http://mamici.flu.ro/divertisment/Fabule/Broasca_si_leul.html	0.23%
11. http://www.personal.psu.edu/rbc4/All_Memos.doc	0.23%
12. http://eee499.gantep.edu.tr/db-completed.php	0.23%
13. http://www.readbag.com/pharmacotherapy-pdf-free-pharm3004-accp-springabs	0.23%
14. https://www.entnet.org/sites/default/files/PQRS%20Measure%20128.pdf	0.23%
15. https://www.science.gov/topicpages/s/social+constructivist+paradigm.html	0.23%
16. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3353127	0.23%
17. https://www.pubfacts.com/search/WHR/2	0.23%
18. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3028965	0.23%
19. https://philpapers.org/s/digital%20economy	0.23%
20. http://www.apta.org/uploadedFiles/APTAorg/Payment/Medicare/Pay_for_Performance/PQRS/201...	0.23%
21. https://www.wikihow.com/Maintain-a-Healthy-Cardiovascular-System	0.23%
22. https://www.cancer.gov/about-cancer/causes-prevention/risk/myths	0.23%
23. https://chrismasterjohnphd.com/2017/08/16/measuring-urine-ph-got-love-working	0.23%
24. https://www.cdc.gov/library/sciclip/issuues/v9issue39.html	0.23%
25. https://www.danepurchasing.com/bids/GetBidDocument/c8d5e148-9338-4a96-9898-02b9c5073db1	0.23%
26. https://www.goodreads.com/author/show/2054109.Jeff_Schweitzer/blog?page=8	0.23%
27. https://quizlet.com/173718530/nhm-311-parkes-flash-cards	0.23%
28. http://journals.lww.com/jnpt/Documents/PEPv29n3S.pdf	0.23%
29. https://www.deathtodiabetes.com/obesity-lose-weight-and-burn-belly-fat.php	0.23%
30. https://www.aafp.org/afp/2010/0615/p1449.html	0.23%
31. http://www.sciencebuddies.org/science-fair-projects/project-ideas/HumBeh_p014/human-behavior...	0.23%
32. http://www.naked-angel.jp/photobbs.myc/photobbs.cgi?mode=disp&no=2566&page=1050	0.23%
33. https://www.wikihow.com/Prevent-Ischemic-Heart-Disease	0.23%
34. https://www.medicographia.com/2010/07/optimizing-secondary-prevention-treatment-in-stable-cor...	0.23%

35. https://laparoscopy.blogs.com/ee06/sls_annual_meeting_and_endo_expo_2006	0.23%
36. http://gistdigest24ng.com/14-signs-partner-cheating	0.23%
37. https://journals.lww.com/ejanaesthesiology/Pages/ArticleViewer.aspx?year=2018&issue=06000&...	0.23%
38. http://www.readbag.com/www3-saintmarys-files-courier-issues-courier-10spring	0.23%
39. http://www.ncor.org.uk/wp-content/uploads/2017/01/BSO-2015-dissertations-abstracts.docx	0.23%
40. https://www.desertfishes.org/dfc/proceed/1997/DFC_Proc97.html	0.23%
41. https://www.nursingcenter.com/journalarticle?Article_ID=1123716	0.23%
42. https://www.nursingcenter.com/journalarticle?Article_ID=1505534	0.23%
43. http://www.health.gov.on.ca/en/common/ministry/publications/plans/ppar17	0.23%
44. https://www.paperdue.com/topic/program-evaluation-essays/7	0.23%
45. https://www.science.gov/topicpages/d/demonstration+program+project.html	0.23%
46. https://writersestate.wordpress.com/category/english/page/69	0.23%
47. http://mason.gmu.edu/%7Ejosterli/Syllabi/summer02b.htm	0.23%
48. https://www.science.gov/topicpages/t/technology+project+review.html	0.23%
49. https://obgyn.onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0412.2012.01435_2.x	0.23%
50. https://open.library.ubc.ca/collections/24/items/1.0355400	0.23%
51. https://www.science.gov/topicpages/a/adult+privacy+policy.html	0.23%
52. http://www.mustbegeek.com/enable-ssh-in-cisco-ios-router	0.23%
53. https://pharmaceuticalintelligence.com/category/myocardial-ischemia	0.23%
54. http://www.spinaldeformities.com/articles.aspx?q=Spinal-Fractures	0.23%
55. https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.106.171016	0.23%
56. https://philpapers.org/s/Environmental%20semiotics	0.23%
57. https://suufoe18.sched.com/list/descriptions	0.23%
58. https://journals.lww.com/nsca-jscr/Pages/articleviewer.aspx?year=2015&issue=08000&article=00...	0.23%
59. http://www.ci.uc.pt/diglit/DigLitWebPdeProjectos.html	0.23%
60. https://www.nbme.org/research/grantsawarded.html	0.23%
61. https://onlinelibrary.wiley.com/doi/full/10.1111/obr.12149	0.23%
62. http://www.cloudshore.com/archives/2010/Sep/2010-09-19_1.html	0.23%

Excluded as citation or reference Web sources: 86 sources found

1. http://seandelaney.com/wp-content/uploads/2012/11/Problems-in-Teaching-Primary-School-Mathem.	0.87%
2. http://www.bolsamadrid.es/esp/asp/Inversores/Agenda/Calendario.aspx	0.55%
3. https://getmorestrength.org/daily/unloading-the-baggage	0.55%
4. https://taxheal.com/section-71-of-insolvency-and-bankruptcy-code-2016.html	0.55%
5. https://taxheal.com/section-68-of-insolvency-and-bankruptcy-code-2016.html	0.55%
6. http://www.readbag.com/ltfraining-portals-ltfraining-docs-english-sat-vocabulary-lessons-sentence...	0.55%
7. http://presbysmediamission.org/2018/02/14/attitude-towards-immigration-reflects-faith-american-idea..	0.55%
8. http://credo.pro/2014/03/131624	0.55%
9. https://propovidnyk.com.ua/46-subota-j-novomisyachchya-1-11-prava-ta-obov-yazky-knyaziv-12-18...	0.55%
10. https://www.treefrogcreative.ca/category/carbon-climate-bioenergy	0.55%
11. https://propovidnyk.com.ua/desyat-zapovidej-v	0.55%
12. http://novzak.uz.ua/news/zakarpatici-proyihali-tri-sotni-kilometriv	0.55%
13. http://www.educa.madrid.org/web/cp.soledad.cubas/nia%20por%20cursos.pdf	0.55%
14. https://nagalandpage.com/skill-building	0.55%
15. https://getmorestrength.org/daily/giving-thanks	0.55%

 Similarity

 Similarity from a chosen source

 Possible character replacement

 Citation

 References

16. http://klavertjevierviel.be/images/Kalender%20Viersel%202018-2019.pdf	0.55%
17. https://propovidnyk.com.ua/litaniya-do-presvyatogo-sertsya-isusa	0.55%
18. https://taxheal.com/section-74-of-insolvency-and-bankruptcy-code-2016.html	0.55%
19. http://credo.pro/2012/03/60388	0.55%
20. https://taxheal.com/companies-incorporation-rules-2014.html	0.55%
21. http://oblrada.odessa.gov.ua/odeska-oblast/byudzheth-odeskoyi-oblasti-na-2016-rik	0.55%
22. https://www.cbs-annedevries.nl/pdf/kalender-ouders-2017-2018.compressed.pdf	0.55%
23. https://propovidnyk.com.ua/o-yatsek-salij-chetverta-zapovid-shanuj-bat-ka-tvogo-matir-tvoyu	0.55%
24. https://klasvanjufsofie.files.wordpress.com/2016/07/maandkalender-zonder-vakanties.pdf	0.55%
25. https://www.physicsclassroom.com/reviews/Waves/Waves-Review-Answers	0.55%
26. https://www.booking.com/hotel/hk/attitude-on-granville.html	0.55%
27. https://panmacmillan.azureedge.net/pml/panmacmillancorporatesite/media/pdf-downloads/free-bu...	0.55%
28. http://oblrada.odessa.gov.ua/odeska-oblast/oblasnyj-byudzheth	0.55%
29. https://taxheal.com/rbi-issued-revised-framework-for-external-commercial-borrowings.html	0.55%
30. https://agnietenschool.nl/wp-content/uploads/sites/11/2018/09/jaarkalender-20182019.pdf	0.55%
31. http://www.est.uc3m.es/esp/nueva_docencia/leganes/ing_telecomunicacion/metodos_mejora_ca...	0.35%
32. http://www.atmseminar.org/seminarContent/seminar8/presentations/pr_152_NSTFO.pdf	0.35%
33. http://www.rb-safeplaceforyouth.com/wp-content/uploads/2017/03/2015CADReport.pdf	0.35%
34. http://www.vollman.com/pdf/PowerofOne102215handout.pdf	0.32%
35. http://semerwp.saluddigital.net/wp-content/uploads/2010/09/Diabetes.pdf	0.32%
36. http://www.esi2.us.es/~vargas/docencia/tca/_examen/P2_0405_C1_C7.pdf	0.32%
37. https://arxiv.org/pdf/1610.06332v1	0.32%
38. https://www.icn.ch/sites/default/files/inline-files/Kathleen%20M%20Vollman.pdf	0.32%
39. https://www.gbiosciences.com/image/pdfs/protocol/BE-405_protocol.pdf	0.32%
40. https://www.dallasfed.org/assets/documents/research/swe/2002/swe0204c.pdf	0.32%
41. https://chemistry.sites.clemson.edu/organic/Labs/2270Docs/Distillation.pdf	0.32%
42. http://cemc.uwaterloo.ca/events/mathcircles/2012-13/Winter/Junior6_Mar5.pdf	0.32%
43. https://link.springer.com/article/10.1186/s13690-018-0269-1	0.32%
44. http://www.cengage.com/resource_uploads/downloads/142406872X_172095.pdf	0.32%
45. https://lc.hkbu.edu.hk/sall/english/doc/Introduction_Steps_writing_graph_description.pdf	0.32%
46. https://staffweb.psdschools.org/bkatz/OnlineMaterials/LSci/Res/Data.pdf	0.32%
47. https://publications.polymtl.ca/1333/1/2013_AliJemmali.pdf	0.32%
48. https://www.chegg.com/homework-help/questions-and-answers/instructions-answer-questions-1-6..	0.32%
49. http://www.nrmp.org/wp-content/uploads/2014/09/Charting-Outcomes-2014-Final.pdf	0.32%
50. https://www.slideshare.net/InfazAhmed/combined-presentation-disprin-v7	0.32%
51. http://ww3.capsim.com/modules/GIA/files/1_0/0/CapsimCore/EN/PDF/UserGuide.pdf	0.32%
52. http://www.d.umn.edu/~rregal/documents/Mingqian/Stat3411_4.1.pdf	0.32%
53. https://arxiv.org/pdf/1705.09985.pdf	0.32%
54. https://docslide.net/engineering/performance-evaluation-of-computationally-efficient-energy-detec...	0.32%
55. https://www.slideshare.net/althafhussain1023/how-to-dimension-user-traffic-in-lte	0.32%
56. https://fr.scribd.com/doc/217993908/tws-contextual-factors	0.32%
57. https://core.ac.uk/download/pdf/70615092.pdf	0.32%
58. https://www.imf.org/~media/Files/Publications/CR/2017/cr1781.ashx	0.32%
59. https://www.slideshare.net/Tommy96/data-mining-using-cultural-algorithms-and-regional-schemata	0.32%
60. http://publications.lib.chalmers.se/records/fulltext/139964.pdf	0.32%
61. https://www.slideshare.net/WillyDevNET/deep-learningapproach-105601622	0.26%



Similarity



Similarity from a chosen source



Possible character replacement



Citation



References

62. http://docshare.tips/capstone_5851a581b6d87fa7438b5896.html	0.23%
63. https://www.slideserve.com/tien/corso-de-core-tools	0.23%
64. https://sfbrealty.sfbli.com/MISC/Files%20for%20future%20use/Admin/Operating%20Statement%...	0.23%
65. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=10000RIL.TXT	0.23%
66. http://infohost.nmt.edu/~olegm/382book/math382sp16.pdf	0.23%
67. http://quod.lib.umich.edu/c/cjs/bby6916.0001.001?rgn=main;view=fulltext	0.23%
68. http://image-processing.blogspot.com	0.23%
69. http://www.esi2.us.es/~vargas/docencia/tca_/ejercicios/discreto/ejerciciosDiscr.pdf	0.23%
70. http://www.wrsd.org/cms/lib04/NJ01000230/Centricity/Domain/389/Study_Guide_Workbook.pdf	0.23%
71. http://www.campusoxide.com/skyDocumentView/Performance-Management-in-SAIL/423	0.23%
72. http://www.cengage.com/resource_uploads/downloads/0495096385_64345.pdf	0.23%
73. https://community.ksde.org/LinkClick.aspx?fileticket=n2ZHaEprfs%3D&tabid=6036&mid=14879	0.23%
74. https://www.paho.org/cub/index.php?option=com_docman&view=download&alias=742-pubfinlay-l...	0.23%
75. http://www.onderwijsvaniedereen.be/2016/04/22/geloof-van-leerlingen-uit-katholiek-onderwijs-neem	0.23%
76. https://www.gestiopolis.com/ejemplo-de-un-informe-de-valoracion-de-una-empresa	0.23%
77. http://classes.bus.oregonstate.edu/ba528/master%20copy%20-%20web%20-%20color%20-%20E..	0.23%
78. http://flightcollege.com.ua/library/2%20%D0%95%D0%A1%D0%A2%D0%95%D0%A1%D0%A2%..	0.23%
79. https://www.ruminantia.it/wp-content/uploads/2018/11/CORNELL-NUTRITION-CONFERENCE-20...	0.23%
80. https://www.slideshare.net/niapola/50085926-ntciso8258	0.23%
81. https://circulodeempresarios.org/app/uploads/2019/01/Libro-la-empresa-mediana-2018.pdf	0.23%
82. https://docplayer.es/55104743-El-antoniano-es-una-revista-cientifica-cultural-de-la-universidad-na...	0.23%
83. https://www.science.gov/topicpages/p/perpetual+foreigner+stereotype.html	0.23%
84. http://www.est.uc3m.es/esp/nueva_docencia/leganes/ing_tec_industrial_todas/metodos_estadist...	0.23%
85. http://www.biblecloud.com/nasb/john/2	0.23%
86. http://www.dupont.com/content/dam/dupont/products-and-services/plastics-polymers-and-resins/e...	0.23%

Library sources: 4 sources found

13020029-HARDIAN ADITYA W.docx	0.26%
13020029-HARDIAN ADITYA W (revisi).docx	0.26%
13020150_Andriko Satria Pratama.docx	0.23%
13020068-Nico Varian Nugroho.docx	0.23%

Excluded as citation or reference Library sources: 68 sources found

12.92.0053 - Maria Stefani-14 MARET.docx	0.72%
Tantri Mapro Psi (1).docx	0.72%
Tantri Rev 1 Mapropsi (1).docx	0.72%
GORBIANO 27 Juli.docx	0.61%
natalia desi 13.12.0017.docx	0.55%
14.I2.0146-MICHAEL ADRIAN ISKANDAR-19 MARET.docx	0.55%
Tugas Kecil 2 Anstruk 2 - Helda Anggelina Christanti_16B100.doc	0.55%
Tugas Kecil 2 Anstruk 2 - Satrio (16.B1.0113).doc.doc	0.55%
14.I2.0146-MICHAEL ADRIAN ISKANDAR-22 MARET.docx	0.55%
Tugas Kecil Anstruk 2 kelas 02_02 - Christian Widiatmo (16.B.docx	0.55%
STEPHANIA ELIZA WIJAYA-28 JUNI.docx	0.55%
Fiona 5 April.docx	0.55%

16.B1.0093-DANIEL HARTONO-TUGAS KECIL 02 ANALISA STRUKTUR II.docx	0.55%
TUGAS2_MICHAEL_16.B1.0009.docx.docx	0.55%
LUSIA DYAH-21 MEI.docx	0.55%
Tugas Kecil 2 Anstruk 2 - Rgel Putra S(16.B1.0021).doc.doc	0.55%
RAMSES BESTRO (16B10028).doc.doc	0.55%
Montana Raisya Putri_ 16.b1.0117_Tugas Kecil 2 Anstruk 2.doc.doc	0.55%
tugas pak Ridwan.docx (1).docx	0.55%
Tomas Cahyo P (16.B1.0068) Tugas kecil 2.doc.doc	0.55%
Perbaikan_Jurnal_13020028.pdf	0.55%
13020028.pdf	0.55%
13020028_Perbaikan_19Jan.pdf	0.55%
JOSUA TH PANGGABEAN (16B10127).doc.doc	0.55%
Jurnal_13020028.odt	0.55%
DICKY ARFENDO SILALAH(16.B1.0101).doc.doc	0.55%
13.02.0028-Stevanus Adi Ramli.doc	0.55%
TUGAS KECIL ANSTRUK 2 INAS ATIKAH_16.B1.0041.docx.docx	0.55%
TUGAS KECIL 2 KORESY ARISTARCHUS 16.B1.0040.docx.docx	0.43%
Tugas Kecil 2 Anstruk 2 - Febrian Iddo Valentino (16.B1.0110.doc	0.43%
perbaikan_13020028.pdf	0.43%
14.I1.0006 - KEZIA MELINDA IRAWAN.docx	0.32%
UKI 14.I1.0103.docx	0.32%
UKI 14.I1.0103 UNICHECK.docx	0.32%
Eolia Kristiani 14.I1.0012.docx	0.32%
Eolia Kristiani 14.I1.0012 Rev1.docx	0.32%
14.I1.0071 - Roy Anggoro.docx	0.32%
13 70 0081 Tithan Dianugrahaeni-22 MARET.docx	0.32%
14.I1.0073-Edwin Widjaja-15 FEB.docx	0.32%
CLAUDIA-10 MARET.docx	0.32%
KEVIN ANDERSON 13.30.0239-14 FEB.docx	0.32%
KEVIN ANDERSON 13.30.0239-10 JAN.docx	0.32%
13 70 0081 Tithan Dianugrahaeni-22 MARET.docx	0.32%
13.07.0028-ENGGAR DWI UTAMI.docx	0.32%
Hanna M 22 Maret.docx	0.32%
13 70 0081 Tithan Dianugrahaen-22 MARET.docx	0.32%
KEVIN ANDERSON 13.30.0239-10 JAN.docx	0.32%
KEVIN ANDERSON 13.30.0239- 9 JAN.docx	0.32%
11.92.0080 Aressa Miga Putranto-7 MARET.docx	0.32%
ANALISIS PORTOFOLIO UTS.pdf.pdf	0.29%
15.G1.0040_Stefanus Rio M.pptx.pptx	0.29%
12.70.0120-Ananda Nugroho.docx	0.29%
14110218_Monika Vresti-2 Jan.docx	0.29%
AYU RAFIDA AL HAFIDZ - 12.92.0016-4 JULI.docx	0.26%
AYU RAFIDA AL HAFIDZ - 2 JULI.docx	0.26%
ALDO FEBRIYAN- 16.B1.0026 (tgs kecil 2).docx.docx	0.23%
perbaikan_sebenarnya_13020028.pdf	0.23%
EvaluasiDiriSisForUnikaSmg.pdf	0.23%



Similarity



Similarity from a chosen source



Possible character replacement



Citation



References

Tugas Kecil 2_Brigita Tara_16.B1.0133.docx.docx	0.23%
16548-16546-1-PB.pdf	0.23%
14.I1.0125-Nathania Janica Theola.docx	0.23%
Aditya Surya Prabowo 14.I1.0077.docx	0.23%
CLAUDIA-14 MARET.docx	0.23%
Theresia Marttiana Woro Sugerti-20 MARET.docx	0.23%
THERESIA MARTTIANA-21 MARET.docx	0.23%
Tugas Kecil 2_Patricia Bella_16.B1.0053.docx.docx	0.23%
12.92.0014 - Lusya Dyah Pratiwi-30 MEI.docx	0.23%
Audrey Ardian Pranoto-30 MEI.docx	0.23%



CHAPTER 1

INTRODUCTION

1.1 Background

Health is a condition where the physical body and mental not merely good but also social well-being. And because being social and productive is not an activities that will be done only once but continuously, so health must be maintained properly. Of course, to maintain the health of each individual there are various ways to do it, namely by examination, prevention and treatment. And if it has been infected by an illness, it must be treated immediately. Certainly it requires different cost for each individual.

Based on the explanation above, to find out the amount of medical cost for each individual, a prediction program that can determine the estimated amount of medical cost is needed. To predict these costs, the backpropagation algorithm is used with data from an insurance service. Backpropagation is a method which used in artificial neural networks that can reduce the error rate with adjusting the weight from actual and prediction output. Backpropagation usually used for prediction, prediction is a process to estimate something that will mostly occur based on the past and current data to reduce the error.

The data for this research is obtained from kaggle.com and can be taken only by downloading it with registered account. The website itself is a trusted source for dataset.

1.2 Scope

The limitation of the problem in this study are as follow:

1. The Training data used is 1000
2. The Testing data used is 250
3. The user need to input the number of hidden layer more than 1

4. The user need to input the number of learning rate in range of 0-1
5. The user need to input the number of iteration more than 1
6. Using Min Max Normalization
7. Using Java language programming
8. The output is in the form of CSV file

Base of the background of this project, the problems to be solved are:

1. Can someone's medical costs be predicted?
2. Is backpropagation algorithm suitable for predicting?
3. How much the error rate calculated from backpropagation algorithm to determine medical costs?

1.3 Objective

The purpose of this project is to produce a program that can predict the medical costs needed by each individual with certain factors. With this prediction an insurance company can make a plan for good hospitality and service for their customer and still have profit. A good hospitality means they can make an insurance program that make their customer joining the program is a profitable action because they will spend less or doesn't have to spend anything in their medical expense.

CHAPTER 2

LITERATURE STUDY

In the journal "Prediction of Agricultural Crops Using Artificial Neural Network Approaches" (Sneehal S. Dahikar, Sandeep V. Rode, 2014), Discusses the use of ANN to predict agricultural crops. In this journal backpropagation is used as an algorithm in the hidden network on ANN. The conclusion in this journal is backpropagation namely ANN is a beneficial tool for prediction. Differences in journals related to the results of this study are the results that will be predicted, namely at this time the results of this project are medical costs.

In the journal "Electronic cigarettes: human health effects" (P.Callahan-Lyon, 2014), discusses the use and effects of smoke from electronic cigarettes on human health. The conclusion in this journal whether cigarette aerosol or e-cigarette aerosol contains toxicants that are harmful to human health. The relationship of this journal to the researcher project is because smoking activity can affect health and certainly will affect medical costs as well, so smoking activities are also used as researchers as one of the parameters in the research project.

In the journal "Stereotypes of Aging: Their Effects on the Health of Older Adults" (Rylee A. Dionigi, 2015), discusses the physical and mental functions that are affected by age and also what can be called "good health" by increasing age. The conclusion in this journal is as someone get older their immunity got older as well. It means that their immunity become weaker and that will make them easier to get sick. The relationship between this journal and the researcher project is that age can affect a person's physical and mental health which certainly also affects the medical costs to be spent. Because of that age becomes one of the factors or parameters that will affect medical costs in this project.

In the journal "Risk for chronic diseases with obesity: Health Survey for England 2010" (Helen L. Maclaughlin, Wendy L. Hall, Thomas AB. SANDERS, Iain C. Macdougall, 2014), discuss some diseases that can be caused by obesity or excessive weight, even at greater risk of contracting CKD (Chronic Kidney Disease) or chronic kidney disease. The conclusion in this journal is "Obesity is associated with an increased risk of CKD in a national sample of the UK population, even after adjustment for known CKD risk factors, which may have implications for CKD screening and future national health service planning and delivery". The relationship of this journal to the research project is the level of obesity or commonly measured from BMI (body mass index) is a factor that affects health and can be used as a parameter in this project.

In the journal "The influence of gender and other patient characteristics on health care-seeking behavior: a QUALICOPC study" (Ashley E. Thompson, Yvonne Anisimowicz, Baukje Miedema, William Hogg, Walter P. Wodchis, Kris Aubrey-Bassler, 2016), discuss the differences in the search for different medical treatments for women and men. With these differences, the costs incurred will be different. The conclusion in this journals is "This study confirms the gender differences in health care-seeking behavior advances previous research by exploring in detail the variables predicting differences in health care-seeking behavior for men and women". The relationship of this journal to the research project is that gender differences can affect health and the way of medical care and cause differences in medical costs, and therefore gender can be a parameter for this project.

CHAPTER 3

RESEARCH METHODOLOGY

Steps that will be carried out by researchers include :

1. Problem analysis, in this case the problems found is the difficulty to determine the total medical costs that must be prepared by a person and with those problems, several factors that can support to determine medical cost will be needed.
2. Study journals and books, from the problem that have been found, journal and books that can support the researcher project are sought. The purpose is to support the researcher project by making it the basis for things that the researcher cannot examine due to lack of time and knowledge.
3. Normalization of data, after the data obtained from the source, the data must be converted to the form of csv to make it easier for the program to read. And then some parameters will be normalized to make it able to be calculated in the algorithm.
4. Coding process, the algorithm will be created in the form of java program that will be used to calculate the data. The data is divided into two kind, the training and the testing data.
5. Training data, the data is processed into the algorithm in the java program that already been created by researcher. And the outcome will be used to determine the testing data.
6. Analysis data, the testing of data will be carried out after the training. Which the result will be compared to the original data to find out the error rate.

CHAPTER 4

ANALYSIS AND DESIGN

4.1 Analysis

The problem in this project is the insurance company can't determine how much they will give to a person that have submitted an insurance claim. They can make an insurance program that will not only help reassure the people to join their insurance program but also make the company get profit from it if cost can be predicted beforehand. This program will make an insurance company can predict how much people spend on medical cost and how much they need to cover from an small injury to big injury.

Table 1: Analysis Data Table

Age	Gender	Body Mass Index	Smoker	Medical Cost
18	male	28.31	no	11272
18	male	39.14	no	12890
18	male	26.125	no	1709
18	female	31.92	no	2206
18	female	36.85	no	1630
19	female	34.7	yes	36398
19	female	20.6	no	1732
19	male	34.1	no	1261
19	male	29.07	yes	17353
19	male	20.9	no	1832
19	male	34.8	yes	34780
20	male	33	no	1980
20	female	21.8	yes	20167
20	female	29	no	2457
21	female	25.8	no	2008
22	female	28.05	no	2156
22	male	25.175	no	2046
23	female	36.67	yes	38512
24	male	35.86	no	1987
24	male	28.5	no	3538
25	male	25.74	no	2138

The data shown in the table above is a portion of training data that have been sorted. Can be seen from the table above even if the age is the same their medical cost differ from one to another. It means the cost can't be determined by only one factor. For such data it need a method that can calculate the complex requirement for the result. Complex requirement means they need more than one factor to calculate the medical cost. To calculate the medical cost each factor have a certain weight. The weight is calculated by a neural network that can determine the weight of a factor that affect the medical cost. Backpropagation is a neural network and suitable one to be used as an algorithm to calculate weight of each factor. The data itself can't be calculated by the program as is. The data need to be normalized to numeric for it can be calculated. The data is normalized using min max normalization.

$$\text{norm} = (\text{value} - \text{min}) / (\text{max} - \text{min})$$

value = The value to be normalized

norm = The value after normalized

min = Minimum value of a given attribute

max = Maximum value of a given attribute

The data will be calculated in the program using backpropagation algorithm after the data has been read and normalized. Backpropagation itself had a number of step. The backpropagation algorithm process is as written below :

1. Generate random value in certain range for all weight

w_{ij} = Weight for input index i to hidden layers index j

w_{0j} = Weight for input bias to hidden layers index j

u_j = Weight for hidden layer index j to output

u_0 = Weight for hidden layer bias to output

2. Calculate the neural activation for hidden layer (Forward Propagation)

$$h_{net\ j} = w_{0j} + \sum x_i \cdot w_{ij}$$

$$h_j = 1 / (1 + e^{-h_{net\ j}})$$

$h_{net\ j}$ = neural activation value for hidden layer before transferred

h_j = neural activation value for hidden layer

3. Calculate the neural activation for output layer (Forward Propagation)

$$y_{net} = u_0 + \sum h_j \cdot u_j$$

$$y = 1 / (1 + e^{-y_{net}})$$

y_{net} = neural activation value for output layer before transferred

y = neural activation value for output layer

4. Calculate Δ weight for hidden layers to output

$$\delta_i = (t - y) \cdot y \cdot (1 - y)$$

$$\Delta u_j = \alpha \cdot \delta_k \cdot h_j$$

t = target output

Δu_j = delta weight hidden layer index j to output index k

5. Calculate Δ weight for input to hidden layers

$$\delta_{net} = \delta_k \cdot u_j$$

$$\delta_j = \delta_{net} \cdot h_j \cdot (1 - h_j)$$

$$\Delta w_{ij} = \alpha \cdot \delta_j \cdot x_i$$

Δw_{ij} = delta weight input index i to hidden layer index j

6. Update weight

$$w_{ij} = w_{ij} + \Delta w_{ij}$$

$$u_j = u_j + \Delta u_j$$

9

7. Repeat step 2 – 6 until max iteration has been reached or loss < target error.

$$\text{loss} = (t - k)^2$$

Testing only use (Forward Propagation or Feed Forward) step 2 and 3 is used.

4.2 Desain

This program can be seen from the following diagram :

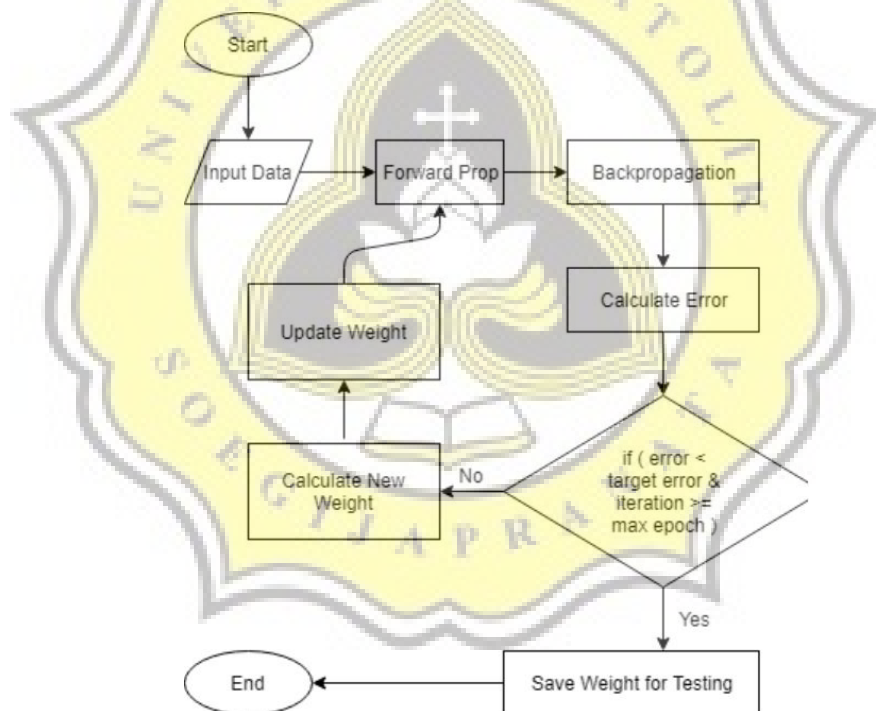
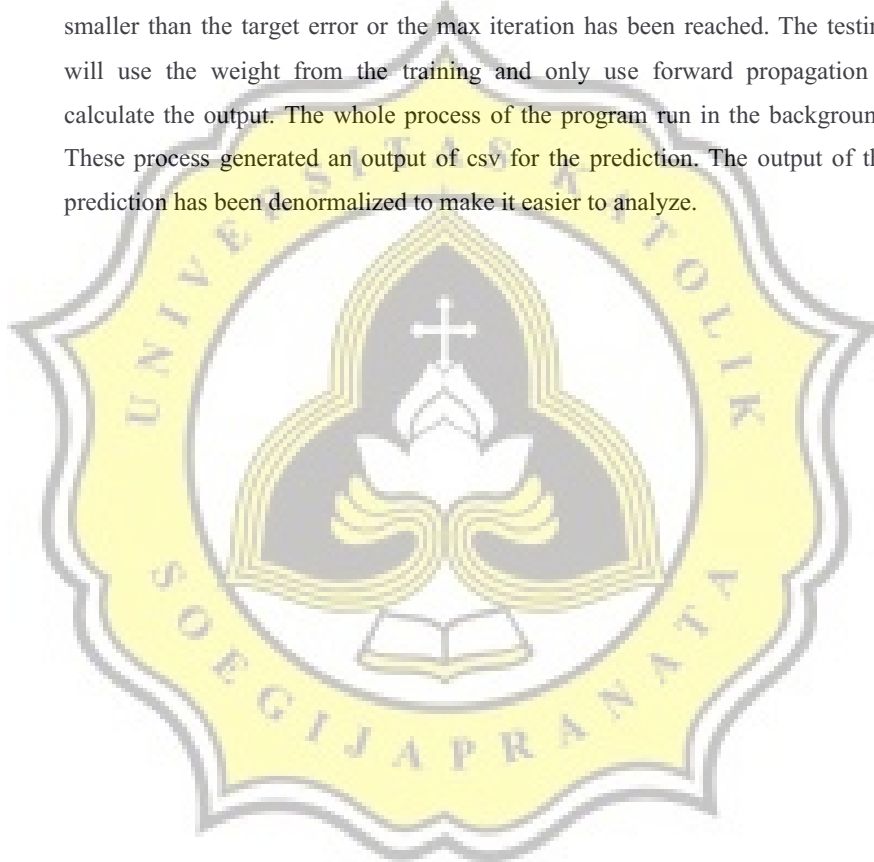


Illustration 1: Backpropagation Flowchart

The program start with the user inputting the data. The data is csv file, hidden layers, learning rate, iteration, and target error. The algorithm begin after the data has been inputted. This algorithm start with training the neural network with feed forwarding the input to a hidden layer using the weight and then to the

10

output. After the output has been calculated, it goes through the backpropagation. In the backpropagation it will calculate the loss using the output and the target, if the loss is higher than the target error it will calculate a new weight and update it. The method repeat in forward propagation and backpropagation until the loss is smaller than the target error or the max iteration has been reached. The testing will use the weight from the training and only use forward propagation to calculate the output. The whole process of the program run in the background. These process generated an output of csv for the prediction. The output of the prediction has been denormalized to make it easier to analyze.



10

CHAPTER 5

IMPLEMENTATION AND TESTING

5.1 Implementation

The implementation of this program start from reading the data in form of CSV and processed in a Java Platform. The first thing the program do after reading the data from CSV file is to give a random weight to the first data of all variable (age, gender, body mass index, smoker) and hidden layers. The random weight used in this research is in range of -0.5 to 0.5.

The algorithm begin with a process called feedforward after the weight has been randomized. In this process, the goal is to calculate the output and compare the output to the target error user have set. The output can't be calculated because the hidden layers still doesn't have a value. In order to give a value for the hidden layer, first the input value and the bias input need to be multiplied with their weight with a formula, after obtaining the hidden layers value the output can be calculated using the same formula to calculate the hidden layers.

The next step begin when the output has been calculated. The next step is to calculate the loss. The loss can be calculated by using the formula to square the difference between output and targeted value ($\text{Math.pow}(\text{output} - \text{targeted_value})$). The current weight can be called the best weight if the loss is smaller than target error or the max iteration has been reached, if the loss is bigger than target error then the iteration continue and the weight will be calculated again using a formula that involve learning rate of the AI.

The new weight is calculated using a formula with a learning. This learning rate determine how fast the neural network changes its mind. The calculation of the new weight begin with calculating the error rate and then calculate the differences weight using the error rate and add the difference weight to the real weight to update it for new weight and input it to the ArrayList, after

FORMULIR SCAN ANTI PLAGIARISME

2,95 %

Nama

: Surya Jatikusuma S

Alamat email

: 15k10017@student.unika.ac.id

Fak. / Prodi

: Ilmu Komputer / Teknik Informatika NIM: 15.K1.0017

berupa(**TESIS, TUGAS AKHIR, PROPOSAL, SKRIPSI, SUMMARY, LAPORAN KERJA PRAKTEK**)

dengan judul

: Medical Cost Prediction Using
Back Propagation Algorithm

Semarang, 25 Januari 2019

Petugas,

Yang Menyerahkan

Dosen Pembimbing

Surya Jatikusuma

Handwritten signature of the supervisor

NB. Laporan hasil scan terlampir

untuk Yang bersangkutan *

PERPUSTAKAAN

the ArrayList for the new hidden layer weight and new variable weight have been calculated replace the current weight with the new one. This step will be repeated with the feedforward until the output is smaller than target loss or the iteration have reached its limit.

The testing process in the program is using only the feedforward process. After the output has been calculated, the loss is also calculated. The loss is needed to calculate the error rate (RMSE).The formula to calculate the error rate is square of (sum of all loss divided size of data). The test can be called a success if the error rate is smaller than the targeted error rate.

5.2 Testing

The test is done with a certain hidden layers, max iteration, learning rate, target error inputed by user and 250 different data from training. Below is result of some testing that used 2000 iteration, 0.9 learning rate, 0.05 target error and inputted hidden layers.

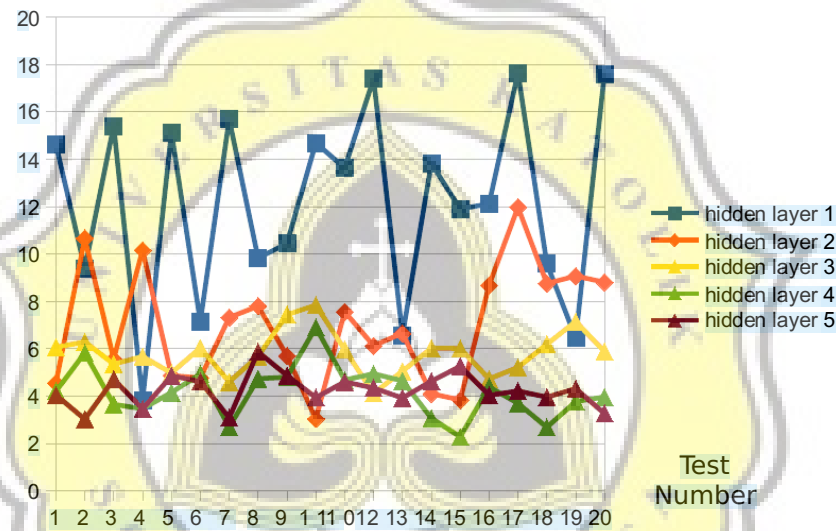
Table 2: Error (RMSE) of testing using different hidden layers in percentage

Test Number	Error Rate (percentage)				
	hidden layer 1	hidden layer 2	hidden layer 3	hidden layer 4	hidden layer 5
1	14.62	4.56	6.07	4.21	4.05
2	9.39	10.67	6.29	5.82	3.03
3	15.41	5.57	5.36	3.65	4.71
4	3.83	10.16	5.67	3.47	3.46
5	15.14	4.87	4.98	4.16	4.86
6	7.14	4.77	6.04	4.94	4.62
7	15.69	7.32	4.6	2.7	3.09
8	9.83	7.8	5.68	4.74	5.87
9	10.44	5.71	7.44	4.8	4.87
10	14.69	3.05	7.84	6.9	3.94
11	13.62	7.56	5.97	4.69	4.61
12	17.4	6.13	4.12	4.95	4.35
13	6.56	6.66	5.05	4.65	3.92
14	13.82	4.1	6.03	3.07	4.63
15	11.9	3.84	6.01	2.28	5.26

13

16	12.12	8.68	4.73	4.39	4.04
17	17.65	11.97	5.21	3.7	4.22
18	9.6	8.76	6.19	2.69	3.96
19	6.45	9.07	7.15	3.77	4.32
20	17.57	8.81	5.89	3.97	3.27

Error Rate
(Percentage)



Can be seen from table and graph above using more hidden layer will have more accurate result. The test with 1 hidden layer had 19 RMSE that exceed the target error (5%) and the graph is not stable. The test with 4 hidden layer had 2 RMSE that exceed the target error. The test with 4 hidden layer had the same total of RMSE that exceed the target error with the test with 5 hidden layer but, the difference between min and max error rate is more bigger than the test with 5 hidden layer. The test with 5 hidden layer had 2 RMSE that exceed the target error. The learning rate set at 0.9 is because the total amount of training data in this project is small, around 1000 data. If the learning rate smaller, it means the neural network will change its mind slower and if it is higher, it means the neural

14

network will change its mind faster. The iteration is adjusting the hidden layer and the learning. If the iteration is smaller the iteration need to be higher because the neural network will change its mind slower and need more loop to get to the targeted error.

Table 3: The differences between real and predicted charges

Age	Gender	BMI	Smoker	Real Charges	Predicted Charges	Differences
33	male	22.705	no	21984	22114.61	130.61
32	male	28.88	no	3867	3923.1	56.1
31	female	25.74	no	3757	3969.51	212.51
46	female	33.44	no	8241	8233.98	7.02
37	female	27.74	no	7282	7247.05	34.95
37	male	29.83	no	6406	6326.8	79.2
60	female	25.84	no	28923	28944.97	21.97
25	male	26.22	no	2721	2729.63	8.63
62	female	26.29	yes	27809	27973.05	164.05
23	male	34.4	no	1827	2174.46	347.46
56	female	39.82	no	11091	11014.18	76.82
27	male	42.13	yes	39612	39791.67	179.67
19	male	24.6	no	1837	1960.23	123.23
52	female	30.78	no	10797	10748.34	48.66
23	male	23.845	no	2395	2431.73	36.73
56	male	40.3	no	10602	10528.28	73.72
30	male	35.3	yes	36837	36620.89	216.11
60	female	36.005	no	13229	13173.4	55.6
30	female	32.4	no	4150	4226.96	76.96

The table above is a portion of prediction output and the differences between the prediction and real charges. The average of the differences between predicted and real charges is 128.174. The prediction obtained from testing using 5 hidden layers, 2000 iteration, 0.9 learning rate, 0.05 target error and got a 4.15% error (RMSE).

CHAPTER 6

CONCLUSION

The conclusions of this project are : Learning from the past journal and from testing the project itself conclude that backpropagation can be used for prediction. The count for the success exceed ninety percent, and because of that backpropagation can be said suitable to predict medical cost. All the input from user (hidden layers, learning rate and iteration) affect the outcome of prediction in certain ways. The minimum of hidden layers needed is around the number of the parameter, and the higher the hidden layers is, the more accurate the prediction will become but it will take more time. Learning rate is how fast the AI will learn if it is set at 1, the AI will change its mind faster, if it is set 0.1 or lower it will change its mind slower and will be more accurate. The slower the neural network change its mind it will need more iteration to make it more accurate.

This project can be improved in the future using the suggestion below :

1. More data to make it more accurate
2. More parameter will make it more accurate to but it need to have a connection with the result or prediction the research made