

Doc vs Internet

81.97% Originality	18.03% Similarity	172 Sources
--------------------	-------------------	-------------

Web sources: 170 sources found

1. https://docplayer.info/30807454-Penerapan-data-mining-untuk-rekomendasi-beasiswa-pada-sma-m...	3.02%
2. http://eprints.dinus.ac.id/13201/1/jurnal_13671.pdf	2.58%
3. http://library.binus.ac.id/eColls/eThesiscoll/Bab2DOC/2012-1-00007-SI%20Bab2001.doc	2.49%
4. http://lppm.bsi.ac.id/SNIT2015/BidangA/A32_197-202_2015-SNIT_Harry%20Dhika_Kajian%20Kom...	2.23%
5. https://docplayer.info/73439454-Penerapan-data-mining-untuk-mengklasifikasikan-penyakit-berdas...	2.14%
6. http://ejournal.nusamandiri.ac.id/ejurnal/index.php/techno/article/download/266/226	1.98%
7. https://docplayer.info/31750651-Implementasi-algoritma-naive-bayes-untuk-memprediksi-penjurusa...	1.6%
8. http://elib.unikom.ac.id/files/disk1/738/jbptunikompp-gdl-danangaris-36853-10-20.unik-a.pdf	1.54%
9. https://ejournal.unsrat.ac.id/index.php/informatika/article/download/17652/17172	1.35%
10. https://github.com/HarunUmar/naive_bayes	1.35%
11. http://yusifisiahaan.blogspot.com	1.35%
12. http://jurnal.fikom.umi.ac.id/index.php/ILKOM/article/download/116/73	1.35%
13. http://achmatim.net/_downloads/publikasi/pengaduan-jumardi-solichin-telematika.pdf	1.23%
14. https://docplayer.info/47366116-Decision-tree-berbasis-algoritma-untuk-pengambilan-keputusan.h...	1.2%
15. http://publication.gunadarma.ac.id/bitstream/123456789/1359/1/50407997.pdf	1.16%
16. http://eprints.dinus.ac.id/12380/1/jurnal_12309.pdf	1.16%
17. https://www.scribd.com/document/389418123/14071998-pdf	1.16%
18. https://docplayer.info/349734-Data-mining-untuk-menganalisa-prediksi-mahasiswa-berpotensi-non...	1.07%
19. https://core.ac.uk/download/pdf/35379064.pdf	1.01%
20. https://docplayer.info/33315965-Implementasi-data-mining-dengan-algoritma-c4-5-untuk-mempred...	0.91%
21. http://achmatim.net/_downloads/publikasi/008.pdf	0.91%
22. https://docplayer.info/47728694-Penerapank-nearest-neighbor-berbasis-genetic-algorithm-untukpe...	0.91%
23. http://riset.budiluhur.ac.id/wp-content/uploads/2013/10/PngmbngnAplikasiTextRecognitionDgnNeu...	0.85%
24. http://umardanny.com/download/0911510238_jurnal.pdf	0.85%
25. https://id.123dok.com/document/9yn0g61q-analisis-gray-level-difference-method-dan-metode-naiv...	0.82%
26. http://belajarcoder.blogspot.com/2016/07/algoritma-c45.html	0.79%
27. http://riset.budiluhur.ac.id/wp-content/uploads/2013/10/VOIPutkKomunikasiAntarKary_Painem.pdf	0.76%
28. http://eprints.dinus.ac.id/18272/2/jurnal_17841.pdf	0.76%
29. http://www.myowndigitalibrary.com/ind.html	0.72%
30. https://candradmy.wordpress.com/page/4	0.69%
31. http://chyonanda.blogspot.com	0.69%
32. http://arifdlh.blogspot.com/2016	0.69%
33. http://fti.budiluhur.ac.id/wp-content/uploads/2014/09/PANDUAN-PENULISAN-KKP-Gasal-201420...	0.69%
34. https://id.123dok.com/document/7qv4eorq-rancang-bangun-sistem-intelijensia-bisnis-untuk-agroin...	0.69%

35. https://seanwerick.com	0.66%
36. https://github.com/conorbmurphy/galvanizereference	0.66%
37. https://advanceddataanalytics.net/whatis-s	0.66%
38. http://catalog.northeastern.edu/course-descriptions/edu	0.66%
39. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4701479	0.66%
40. http://www.academia.edu/Documents/in/Jaringan_Syaraf_Tiruan	0.66%
41. http://docshare.tips/dmcs2009workshopproceedings4_587b480eb6d87f2bbc8b5116.html	0.66%
42. http://budiluhur.academia.edu/AchmadSolichin	0.66%
43. http://catalog.northeastern.edu/undergraduate/university-academics/education	0.66%
44. https://github.com/conorbmurphy/galvanizereference/blob/master/README.md	0.66%
45. http://ejournal.nusamandiri.ac.id/ejournal/index.php/pilar/article/download/340/272	0.66%
46. http://www.umuc.edu/academic-programs/masters-degrees/data-analytics.cfm?customel_dataPa...	0.66%
47. http://umuc.edu/academic-programs/masters-degrees/data-analytics.cfm	0.66%
48. http://dml.cs.byu.edu/~cgc/docs/ml_dm_tools/Reading/PlanningDMPProjects	0.66%
49. https://docplayer.net/3350061-Pattern-mining-a-key-to-improve-network-usage-understanding.html	0.66%
50. https://www.slideshare.net/Tommy96/data-mining-applications-4035391	0.66%
51. https://docplayer.info/33486486-Data-mining-classification-untuk-prediksi-lama-masa-studi-mahas...	0.66%
52. https://en.wikipedia.org/wiki/Data_mining	0.66%
53. https://elhilorojodelmarketing.wordpress.com/2013/08/27/la-leyenda-de-los-gatos-carey-ii-desentra...	0.66%
54. http://www.acheronanalytics.com/acheron-blog/previous/2	0.66%
55. https://www.gsaadvantage.gov/ref_text/GS35F0486W/0TKL50.3NTAV2_GS-35F-0486W_PROTE...	0.66%
56. https://docplayer.info/31622739-Penerapan-algoritma-c4-5-berbasis-adaboost-untuk-prediksi-peny...	0.66%
57. https://www.ibm.com/developerworks/community/blogs/ibm-bi-capabilities?sortby=2&maxresults=...	0.66%
58. https://vdocuments.site/revista-campus-virtuales-01-iv.html	0.66%
59. https://quizlet.com/161934705/dbw-final-flash-cards	0.66%
60. http://webdev.sps.northwestern.edu/program-areas/graduate/data-science	0.66%
61. https://sites.google.com/site/informationandcomputerscience/database-search-results-for-comput...	0.66%
62. https://semanticcommunity.info/data_science/nsf_funding_for_big_data_and_data_science/nsf_gra...	0.66%
63. https://docplayer.net/19895767-Data-mining-case-studies.html	0.66%
64. https://docplayer.info/34834392-Analisis-teknik-data-mining-algoritma-c4-5-dan-k-nearest-neighbo...	0.63%
65. https://fmipa.unmul.ac.id/files/docs/JIM_METODE%20NAIVE%20BAYES%20UNTUK%20PENE...	0.63%
66. https://fjrhdhp.files.wordpress.com/2014/12/paper001.pdf	0.63%
67. http://riset.budiluhur.ac.id/wp-content/uploads/2013/01/Bit-Vol-9-No-2-02.pdf	0.6%
68. http://riset.budiluhur.ac.id/wp-content/uploads/2014/01/KajianKerangkaKerjaPemringktnEGov_Pip...	0.6%
69. https://dianrdntelkomuniversity.wordpress.com/2016/10/23/analisis-model-prediksi-elektabilitas-ca...	0.57%
70. https://www.slideshare.net/msyani/pengamanan-data-dengan-menggunakan-algoritma-kriptografi-..	0.57%
71. http://jitter.widyatama.ac.id/index.php/Selidik2016/article/download/117/95	0.57%
72. http://jurnal.umrah.ac.id/wp-content/uploads/gravity_forms/1-ec61c9cb232a03a96d0947c6478e52...	0.53%
73. https://achmatim.net/publikasi	0.53%
74. http://achmatim.net/publikasi	0.53%
75. http://riset.budiluhur.ac.id/wp-content/uploads/2013/10/AnalisaRancangSIAntrian_RizkyTahara_G...	0.5%
76. http://dspace.upsurabaya.ac.id:8080/xmlui/bitstream/handle/123456789/61/Penerapan%20Algor...	0.5%
77. https://rully6092.wordpress.com/data-mining-data-warehouse	0.5%
78. http://eprints.dinus.ac.id/13239/1/jurnal_13789.pdf	0.5%
79. http://www.testbanks01.com/index.php/2016/11	0.47%
80. https://github.com/CognitiveBuilder/HelloCognitiveWorld/blob/master/code/textbooks.csv	0.47%

 Similarity

 Similarity from a chosen source

 Possible character replacement

 Citation

 References

81. http://beyoureself-kepis.blogspot.com/search/label/programs	0.41%
82. https://www.scribd.com/document/324016766/Paper-Review-Natural-Language-Processing	0.41%
83. http://repository.telkomuniversity.ac.id/pustaka/files/105570/jurnal_eproc/penerapan-teknik-data-m...	0.41%
84. http://docshare.tips/librocas-en-iberoamerica_5750bb4db6d87f5e5a8b48a6.html	0.38%
85. https://machinelearningmastery.com/tutorial-to-implement-k-nearest-neighbors-in-python-from-scr...	0.38%
86. http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/1446/505	0.38%
87. http://jurnaleeccis.ub.ac.id/index.php/eeccis/article/download/204/176	0.38%
88. https://docplayer.es/790118-Colegio-de-postgraduados-data-warehouse-y-mineria-de-datos-como-	0.38%
89. http://achmatim.net/_downloads/publikasi/051-cahya-solichin.pdf	0.38%
90. https://docplayer.info/47728698-Penentuan-penilaian-kredit-menggunakan-metode-naive-bayes-be	0.38%
91. http://ejurnal.net/portal/index.php/ticom/article/view/317/277	0.38%
92. https://core.ac.uk/download/pdf/32453091.pdf	0.38%
93. https://docplayer.net/33476575-Data-analytics-the-first-international-conference-on-data-analytics...	0.35%
94. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0031-2	0.35%
95. http://fportfolio.petra.ac.id/user_files/04-021/Makalah%20Maturity%20UPN.pdf	0.31%
96. http://tugaskuliah-sabanamulia.blogspot.com	0.31%
97. https://jfin-swufe.springeropen.com/track/pdf/10.1186/s40854-016-0029-6?site=jfin-swufe.springe...	0.28%
98. https://text-id.123dok.com/document/1y9m7gwq-klasifikasi-nasabah-kartu-kredit-menggunakan-a...	0.28%
99. http://riset.budiluhur.ac.id/wp-content/uploads/2012/02/BIT-08-2-08.pdf	0.28%
100. https://www.scribd.com/document/322901846/Optimasi-Pembelian-Oleh-Oleh-Di-Kota-Yogyakarta...	0.28%
101. https://www.slideshare.net/agungsulistyankkpsistem-penunjang-keputusan-untuk-menentukan-a...	0.28%
102. https://ojs.unikom.ac.id/index.php/komputa/article/download/48/46	0.28%
103. http://pascasarjana.budiluhur.ac.id/wp-content/uploads/2013/02/Lusi_Nazori_TM-Vol4-No12.pdf	0.28%
104. http://is.its.ac.id/pubs/oajis/index.php/file/download_file/452	0.28%
105. https://docplayer.info/233496-Sistem-pakar-penentuan-bahan-pangan-yang-tepat-untuk-pemenuh	0.28%
106. http://is.its.ac.id/pubs/oajis/index.php/file/download/5/452	0.28%
107. https://docplayer.info/46525521-Perancangan-aplikasi-sistem-pakar-untuk-mendiagnosa-hama-d...	0.28%
108. http://februari-16th.blogspot.com/2013/07/rangkuman-jurnal-jurnal-digital-image.html	0.25%
109. https://tatangmanguny.wordpress.com/2010/03/20/signifikansi-hasil-penelitian	0.25%
110. http://www.ijccts.org/books_pdf_dwd/Predicting%20the%20Number%20of%20Blood%20Donors...	0.25%
111. http://www.testbanks01.com/index.php/category/test-bank/page/29	0.25%
112. http://jtiik.ub.ac.id/index.php/jtiik/article/download/204/pdf	0.25%
113. https://flylib.com/books/en/2.706.1/using_subclasses_and_inheritance.html	0.25%
114. https://flylib.com/books/en/2.253.1/a1_operator_precedence.html	0.25%
115. http://www.ijemr.net/DOC/DataMiningTechniquesApplicationsAndScope(358-365).pdf	0.25%
116. https://flylib.com/books/en/2.345.1.30/1	0.25%
117. https://flylib.com/books/en/2.265.1/basic_instructions.html	0.25%
118. https://docplayer.info/46599338-Penerapan-jst-backpropagation-untuk-prediksi-curah-hujan-studi..	0.25%
119. http://www.redalyc.org/html/2833/283321909007/index.html	0.25%
120. http://download.portalgaruda.org/article.php?article=14831&val=979&title=ANALISIS%20DAN%...	0.25%
121. https://flylib.com/books/en/2.204.1.10/1	0.25%
122. https://text-id.123dok.com/document/myjv5myl-analisis-akurasi-algoritma-naive-bayes-pada-klas...	0.25%
123. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0015-2	0.25%
124. http://ichbinrochmah.blogspot.com	0.25%
125. http://jurnalwacana.psikologi.fk.uns.ac.id/index.php/wacana/article/view/38/38	0.25%
126. https://haluwin.files.wordpress.com/2013/10/aplikasi-penyusunan-jadwal-dengan-algoritma-genet..	0.25%

 Similarity

 Similarity from a chosen source

 Possible character replacement

 Citation

 References

127. https://haifengl.github.io/bigdata	0.25%
128. http://mx.123dok.com/document/myjme6my-analisis-de-la-produccion-cientifica-de-mexico-en-e...	0.25%
129. http://ichbinrochmah.blogspot.com/2014/06/software-testing.html	0.25%
130. https://flylib.com/books/en/2.647.1/culture_and_religion.html	0.25%
131. http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/download/27/25	0.25%
132. https://flylib.com/books/en/2.857.1.15/1	0.25%
133. https://flylib.com/books/en/2.28.1/restructuring.html	0.25%
134. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4766610	0.25%
135. http://journals.ums.ac.id/index.php/khif/issue/download/262/11	0.25%
136. https://www.ijeat.org/download/volume-1-issue-3	0.25%
137. http://ojs.amikom.ac.id/index.php/semnasteknomedia/article/download/1764/1490	0.25%
138. https://www.slideshare.net/AlexanderDecker/analyzing-undergraduate-students-performance-in-v...	0.25%
139. https://ti.ukdw.ac.id/ojs/index.php/informatika/article/download/331/pdf	0.25%
140. http://student-saver.blogspot.com/2015/12/list-of-smtb-books-part-1.html	0.25%
141. https://flylib.com/books/en/2.930.1.46/1	0.25%
142. https://flylib.com/books/en/2.857.1.13/1	0.25%
143. http://dekajus.blogspot.com/2011/04/pengaruh-struktur-aktiva-ukuran_07.html	0.25%
144. https://flylib.com/books/en/3.328.1.14/1	0.25%
145. http://morphinpharmacy.blogspot.com/2013/06	0.25%
146. http://is.its.ac.id/pubs/oajis/index.php/file/download_file/1582	0.25%
147. http://eprints.dinus.ac.id/12192/1/jurnal_12100.pdf	0.25%
148. http://sisteminfomasi.blogspot.com/2016/04	0.25%
149. http://download.portalgaruda.org/article.php?article=14831&val=979	0.25%
150. http://if.binadarma.ac.id/sipi/jurnal/Jurnal-JURNAL%20DENY%20WAHYUDI.pdf	0.25%
151. http://testbanks01.com/index.php/author/admin/page/36	0.25%
152. https://flylib.com/books/en/2.159.1.12/1	0.25%
153. http://www.ala.org/awardsgrants/awards/370/all_years	0.25%
154. https://flylib.com/books/en/2.159.1.17/1	0.25%
155. https://flylib.com/books/en/2.217.1.59/1	0.25%
156. https://patents.google.com/patent/US20170061331A1/en	0.25%
157. http://testbanks01.com/index.php/tag/sm/page/27	0.25%
158. https://flylib.com/books/en/2.145.1/infosec_acceptable_use_policy.html	0.25%
159. http://www.pur-plso.unsri.ac.id/userfiles/21_%20Kartika.pdf	0.25%
160. http://repository.uksw.edu/bitstream/123456789/13708/1/T1_672013221_Full%20text.pdf	0.25%
161. https://flylib.com/books/en/2.345.1.71/1	0.25%
162. https://flylib.com/books/en/1.48.1.57/1	0.25%
163. https://flylib.com/books/en/1.36.1/buffer_overflows.html	0.25%
164. https://flylib.com/books/en/3.194.1.48/1	0.25%
165. https://docplayer.es/1751053-La-mineria-de-datos-como-herramienta-para-la-toma-de-decisiones..	0.25%
166. http://ejournal.net/portal/index.php/ticom/article/view/301/261	0.25%
167. http://portal.ejournal.net/index.php/ticom/article/view/315/275	0.25%
168. https://core.ac.uk/download/pdf/35370823.pdf	0.25%
169. http://februari-16th.blogspot.com/2013	0.25%
170. http://student-saver.blogspot.com/2015/12/list-of-smtb-books-part-2.html	0.25%

Web omitted sources: 2 sources found

 Similarity

 Similarity from a chosen source

 Possible character replacement

 Citation

 References

-
- | | |
|--|-------|
| 1. http://eprints.dinus.ac.id/13487/1/jurnal_14148.pdf | 5.63% |
| 2. https://ejournal.fikom-unasman.ac.id/index.php/jikom/article/download/26/17 | 5.57% |

 Similarity

 Similarity from a chosen source

 Possible character replacement

 Citation

 References

ANALISIS DALAM MENENTUKAN PREDIKSI KEBERHASILAN PENAWARAN KREDIT BAGI PENSIUNAN PEGAWAI NEGERI SIPIL (STUDI KASUS: PT BANK XYZ)

Yohanes Yudhi Perkasa Palendeng¹, Achmad Solichin²

¹Program Studi Magister Ilmu Komputer, Program Pascasarjana, Universitas Budi Luhur

²Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur
Jl. Raya Ciledug, Petukangan Utara, Kebayoran Lama, Jakarta Selatan 12260

Telp. (021) 5853753, Fax. (021) 5866369

E-mail : yohanes.yudhi@gmail.com¹, achmad.solichin@budiluhur.ac.id²

ABSTRACT

Utilization of data mining in marketing strategies of banking is very effective. Customer segmentation of retirees who will be offered credit products is one of the process is done in the marketing strategy of banking. To support the results of the success rate of sales force to market the product in his role of banking credit that the process requires customer data retired, then support data mining was very instrumental in the classification of potential borrowers of the bank so that it can predict the success rate in the marketing of those services. Based on the mapping of research about the customer data mining support retirees so it was decided to research this appealing 2 methods of classification data mining, that Naive Bayes and C4.5. The results showed that the accuracy of C4.5 algorithm is 81.25%, better than Naive Bayes method of 73.54%. In the meantime, user acceptance testing of the system results in a 90% more user acceptance rate.

Keywords: data mining, Naive Bayes, C4.5, classification, credit prediction

ABSTRAK

Pemanfaatan data mining dalam strategi pemasaran perbankan sangat efektif. Segmentasi nasabah pensiunan yang akan ditawarkan produk kredit merupakan salah satu proses yang dilakukan dalam strategi pemasaran perbankan. Untuk mendukung hasil dari tingkat keberhasilan tenaga pemasaran dalam perannya untuk memasarkan produk kredit perbankan yang prosesnya membutuhkan data-data nasabah pensiunan ini, maka dukungan data mining sangat berperan penting dalam klasifikasi calon nasabah bank sehingga dapat memprediksi tingkat keberhasilan dalam pemasaran produk layanan tersebut. Berdasarkan pemetaan penelitian mengenai dukungan data mining pada nasabah pensiunan maka penelitian ini diputuskan untuk membandingkan 2 metode klasifikasi data mining, yaitu Naive Bayes dan C4.5. Hasil penelitian menunjukkan tingkat akurasi algoritma C4.5 yaitu sebesar 81,25%, lebih baik dibandingkan metode Naive Bayes sebesar 73,54%. Sementara itu, pengujian tingkat penerimaan pengguna terhadap sistem menghasilkan tingkat penerimaan pengguna sebesar 90% lebih.

Kata kunci : data mining, Naive Bayes, C4.5, klasifikasi, prediksi kredit

1. PENDAHULUAN

Kredit adalah cara menjual barang dan atau pinjaman uang dengan pembayaran secara tidak tunai dimana pembayaran ditangguhkan atau diangsur dengan pinjaman sampai batas jumlah tertentu yang diizinkan oleh bank atau badan lain. Salah satu tugas utama dari sebuah lembaga keuangan adalah untuk mengembangkan beberapa set model dan teknik untuk memungkinkan mereka untuk memprediksi kebangkrutan dan untuk menilai kredibilitas pelanggan. Pemasaran sendiri adalah suatu proses tentang pengembangan produk, periklanan, distribusi dan penjualan [2].

Konsep *data mining* sendiri merupakan bagian dari konsep teknologi informasi yang berkaitan dengan data dan informasi. *Data mining* sebagai proses atau teknik pemodelan yang mempergunakan analisis dengan variasi data yang banyak untuk mendapatkan pola dan hubungan diantara variasi data tersebut [1]. Kehadiran *data mining* dilatarbelakangi dengan

adanya masalah data explosion atau ledakan data yang dialami oleh banyak organisasi yang telah mengumpulkan data sekian tahun lamanya (data pembelian, data penjualan, data nasabah, data transaksi, dan data-data lainnya) [17]. Data pembelian, data penjualan, data nasabah, data transaksi, dan data lain dalam perusahaan diperoleh oleh perusahaan dari kegiatan operasional yang dilakukan perusahaan. Perusahaan-perusahaan besar dalam berbagai industri, misalnya manufaktur, perbankan, dan retail atau transaksi kartu kredit dari sebuah bank dalam seharusnya tentu akan menghasilkan banyak data dan di dalam data tersebut dapat terkandung informasi yang penting bagi perusahaan. Adanya banyak data dalam perusahaan inilah yang disebut dengan data explosion atau ledakan data.

Dapat dibayangkan besarnya ukuran data yang didapatkan jika nanti proses ini telah berjalan beberapa tahun dan sangat rugi bagi perusahaan jika dari data-data tersebut tidak didapatkan sebuah informasi. Berdasarkan uraian diatas serta mengingat begitu pentingnya *data mining* terhadap operasional suatu organisasi, khususnya sektor perbankan demi

mengoptimalkan penggunaan dana pemasaran, maka sudah seharusnya manajemen menerapkan konsep *data mining* dan melihat manfaatnya untuk memutuskan serta menawarkan kredit bagi para pensiunan pegawai negeri sipil.

Berdasarkan latar belakang yang ada, maka dapat diidentifikasi beberapa masalah antara lain (1) data yang ada masih berupa data mentah, yang masih harus diklasifikasi; (2) penawaran produk kredit yang tidak tepat sasaran; (3) penggunaan cara-cara manual dalam memutuskan serta menawarkan kredit bagi para pensiunan mengakibatkan proses penawaran yang relatif lama; dan (4) kurangnya pemanfaatan data-data transaksi serta data penyaluran pensiun untuk mendapatkan informasi.

Pada penelitian ini ditetapkan beberapa batasan penelitian. Pada penelitian ini data yang digunakan hanya data nasabah pensiunan PNS PT. Bank XYZ sebagai mitra bayar dari PT. Taspen (Persero) dan data pinjaman pensiunan dari tahun 2014 sampai dengan tahun 2015 yang akan dilakukan seleksi. Selain itu penelitian ini hanya berfokus untuk memprediksi tingkat keberhasilan penawaran pemberian kredit. Penelitian ini akan membandingkan 2 buah metode yaitu Naïve Bayes dan C4.5 dan akan dipilih salah satu metode dengan tingkat akurasi yang paling tinggi.

2. LANDASAN TEORI

2.1. Data Mining

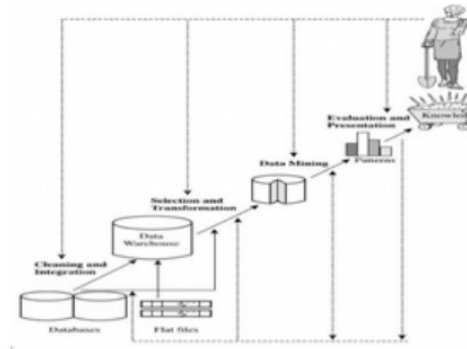
Secara singkat, *data mining* berarti menggali atau menemukan informasi dari sejumlah besar data dan secara luasnya, *data mining* adalah proses menemukan *interesting knowledge* dari sejumlah besar data yang tersimpan dalam database, data warehouse, atau media penyimpanan lainnya [4]. Sedangkan berdasarkan Larose, 2005, *data mining* adalah sebuah proses untuk menemukan hubungan, pola, dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam media penyimpanan, menggunakan teknologi pengenalan pola serta statistika dan matematika [9].

Dengan berdasarkan pengertian diatas, maka dapat dikatakan jika *data mining* adalah sebuah proses mencari informasi dari data yang berjumlah besar dan tersimpan dalam media penyimpanan.

2.2. Metodologi Data Mining

a. Knowledge Discovery from Data

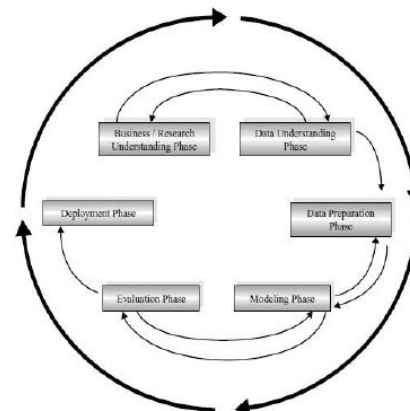
KDD merupakan singkatan dari *Knowledge Discovery from Data*. KDD mulai dikembangkan pada era awal 1990-an. Fayyad pada 1996 menggagas proses model KDD dan menetapkan langkah untuk proyek DM. Pada metodologi KDD terdapat 9 tahap sesuai dengan gambar di bawah [4]



Gambar II- 1. Tahapan Metodologi KDD

b. Cross Industry Standard Process for Data Mining

CRISP-DM merupakan singkatan dari *Cross Industry Standard Process for Data Mining*. CRISP-DM merupakan standarisasi *data mining* yang disusun oleh tiga penggagas *data mining* market. Yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR [9]. Pada metodologi ini dilakukan pembagian siklus untuk proses *data mining* menjadi 6 tahap, dimana ketergantungan antara setiap tahap digambarkan dengan panah. Berikut merupakan gambaran dari metodologi CRISP-DM.



Gambar II- 2. Tahapan Metodologi CRISP-DM [13]

2.3. Algoritma Naïve Bayes

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan "naive" dimana diasumsikan kondisi antar atribut saling bebas. Pada sebuah dataset, setiap baris/dokumen I diasumsikan sebagai vector dari nilai-nilai atribut $\langle x_1, x_2, \dots, x_n \rangle$ dimana tiap nilai-nilai menjadi peninjauan atribut $X_i \in \{c_1, \dots, c_k\}$. Setiap baris mempunyai label kelas $c_i \in \{c_1, c_2, \dots, c_k\}$ sebagai nilai variabel kelas C, sehingga untuk melakukan klasifikasi dapat dihitung nilai

probabilitas $p(C=c_i | X=x_j)$, dikarenakan pada Naïve Bayes diasumsikan setiap atribut saling bebas, maka persamaan yang didapat adalah sebagai berikut:

1. Peluang $p(C=c_i | X=x_j)$, menunjukkan peluang bersyarat atribut X_i dengan nilai x_i diberikan kelas c , dimana dalam Naïve Bayes, kelas C bertipe kualitatif sedangkan atribut X_i dapat bertipe kualitatif ataupun kuantitatif.
2. Ketika atribut X_i bertipe kuantitatif maka peluang $p(C=c_i | X=x_j)$, akan sangat kecil sehingga membuat persamaan peluang tersebut tidak dapat diandalkan untuk permasalahan atribut bertipe kuantitatif. Maka untuk menangani atribut kuantitatif, ada beberapa pendekatan yang dapat digunakan seperti distribusi normal (Gaussian):

$$f = N(X_i; \mu_{c_i}, \sigma_c) = \frac{1}{\sqrt{2\pi}\sigma_c} e^{-\frac{(X_i - \mu_{c_i})^2}{2\sigma_c^2}}$$

Ataupun kernel density estimation (KDE) :

$$f = \frac{1}{n_c} \sum N(X_i; \mu_{c_i}, \sigma_c)$$

2.4. Algoritma C4.5

Quinlan mengemukakan bahwa C4.5 adalah algoritma yang digunakan untuk klasifikasi data yang dapat mengolah data/atribut numerik, algoritma ini dapat mengatasi nilai atribut yang hilang, dan dapat mengatasi data kontinu dan pruning/penyederhanaan [6]. Hasil dari proses klasifikasi berupa aturan yang dapat digunakan untuk memprediksi nilai atribut bertipe diskret dari *record* yang baru. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 dan secara umum digunakan untuk membangun pohon keputusan dengan melakukan tahapan sebagai berikut: pilih atribut sebagai akar (root), buat cabang untuk tiap-tiap nilai, bagi kasus dalam cabang, kemudian ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 yaitu: [9]

1. Mempersiapkan data *training*.
2. Menghitung akar dari pohon. Untuk menghitung nilai entropy digunakan rumus :

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i$$

Keterangan :

S= Himpunan kasus

n = jumlah partisi S

P_i = proporsi S_i terhadap S

3. Kemudian hitung nilai gain menggunakan rumus :

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i)$$

X

Keterangan :

S = Himpunan Kasus

A = Fitur

n = jumlah partisi atribut A

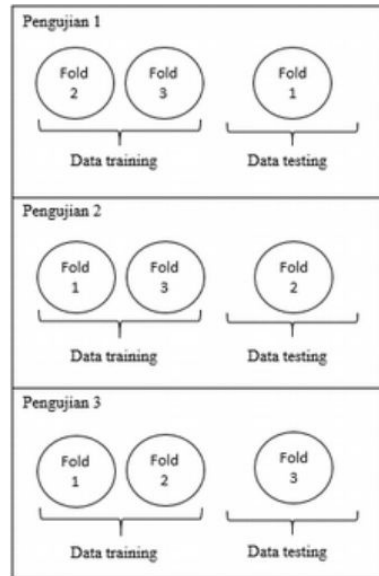
| S_i | = Proporsi S_i terhadap S

|S| = jumlah kasus dalam S

4. Ulangi langkah ke 2 dan langkah ke 3 hingga semua *record* terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut didalam *record* yang dipartisi lagi
 - c. Tidak ada *record* didalam cabang yang kosong.

2.5. K-Fold Cross Validation

Cross-validation adalah metode statistik yang mengevaluasi dan membandingkan algoritma pembelajaran dengan membagi data menjadi dua yaitu data *training* dan data *testing*. Untuk mengilustrasikan metode *k-fold cross validation* ini, dapat dilihat pada gambar dibawah



Gambar II-4. Contoh Ilustrasi 3-fold cross validation

2.6. Hipotesis

Berdasarkan beberapa tinjauan studi serta kerangka konsep yang telah dirancang, maka peneliti dapat menarik hipotesis:

1. Diduga algoritma C4.5 akan mempunyai prediksi akurasi yang lebih tinggi daripada metode Naïve Bayes dalam menentukan prediksi tingkat keberhasilan penawaran pemberian kredit bagi para pensiunan.
2. Dengan adanya sistem diduga akan mempercepat dan mempermudah prediksi penawaran kredit.

3. RANCANGAN SISTEM DAN APLIKASI

3.1. Metode Penelitian

Pada penelitian ini, jenis penelitian yang diambil adalah eksperimen komparatif yaitu dengan membandingkan dua algoritma Naïve Bayes dan C4.5.

3.2. Metode Pemilihan Sampel

Dalam penelitian ini, peneliti mengambil sampel dari populasi jumlah nasabah pensiunan yang ada pada PT. Bank XYZ menggunakan teknik *random sampling* dengan metode *stratified random sampling*.

3.3. Metode Pengumpulan Data

Pada penelitian ini, metode pengumpulan data yang digunakan adalah

1. Observasi

Peneliti melakukan pengamatan langsung di PT. Bank XYZ divisi Rencana Bisnis untuk mengumpulkan data yang

berhubungan dengan penawaran kredit dengan cara mengamati dan mencatat secara sistematis masalah-masalah yang diselidiki dan meneliti secara langsung terhadap objek yang akan diteliti.

2. Wawancara

Peneliti melakukan wawancara tidak terstruktur kepada pihak-pihak yang berkompeten atau berkepentingan dalam menentukan forecast penawaran produk kredit dan yang mengetahui nilai aktual penjualan produk tersebut. Dari hasil wawancara ini diharapkan dapat menambah kelengkapan data yang diperoleh dari hasil pengamatan.

3. Studi Pustaka

Suatu bentuk riset yang menggunakan proses pencarian data dengan cara mencari, membaca buku dan mengolah isi dari beberapa referensi buku atau publikasi lain yang dapat dijadikan acuan dalam pencarian data. Data yang diperoleh dari studi pustaka inilah yang disebut dengan data sekunder, tujuan dari data sekunder ini adalah sebagai landasan teori untuk menganalisa pemecahan masalah di dalam penelitian ini.

3.4. Teknik Analisis

Teknik analisis yang akan dilakukan menggunakan metode-metode *data mining* antara lain Naïve Bayes dan C4.5. Metode-metode tersebut akan digunakan untuk mengolah data nasabah pensiunan PT. Taspen dalam kaitannya dengan penawaran produk kredit di PT. Bank XYZ.

Data akan dibagi menjadi dua set yaitu sebagai data training dan data uji. Hasil pembelajaran masing-masing metode dengan data training akan dibandingkan hasil pengujian menggunakan metode pengujian *10-folds cross validation* untuk mendapatkan nilai-nilai statistik berupa nilai akurasi, *precision*, *recall*, *ROC Curve*. Hasil statistik tersebut akan dibandingkan untuk mendapatkan metode terbaik yang akan diterapkan dalam perancangan prototipe sistem.

3.5. Teknik Pengujian

User Acceptance Testing merupakan pengujian yang dilakukan oleh end-user dimana user tersebut adalah staff/karyawan perusahaan yang langsung berinteraksi dengan sistem dan dilakukan verifikasi apakah fungsi yang ada telah berjalan sesuai dengan kebutuhan/fungsinya [11].

Setelah dilakukan *system testing*, *acceptance testing* menyatakan bahwa sistem software memenuhi persyaratan. *Acceptance testing* merupakan pengujian yang dilakukan oleh pengguna yang menggunakan teknik pengujian black box untuk menguji sistem terhadap spesifikasinya. Pengguna akhir bertanggung jawab untuk memastikan semua fungsionalitas yang relevan telah diuji [10].

4. HASIL DAN PEMBAHASAN

4.1. Confusion Matrix Algoritma C4.5

Dari perhitungan terhadap 6 atribut dengan 480 *record* maka ditemukan 42 *record* diklasifikasikan Approve dan 348 *record* secara benar diklasifikasikan Tdk Approve.

Tabel IV.6. Model Tabel *Confusion Matrix* Algoritma C4.5

Accuracy: 81,25%

	True Tdk Approve	True Approve	Class Precision
Pred. Tdk Approve	348	79	81,50%
Pred. Approve	11	42	79,25%
Class Recall	96,94%	34,71%	

PerformanceVector

```
PerformanceVector:
accuracy: 81.25% +/- 3.49% (mikro: 81.25%)
ConfusionMatrix:
True: Tdk Approve Approve
Tdk Approve: 348 79
Approve: 11 42
precision: 81.50% +/- 15.14% (mikro: 79.25%) (positive class: Approve)
ConfusionMatrix:
True: Tdk Approve Approve
Tdk Approve: 348 79
Approve: 11 42
recall: 34.55% +/- 10.15% (mikro: 34.71%) (positive class: Approve)
ConfusionMatrix:
True: Tdk Approve Approve
Tdk Approve: 348 79
Approve: 11 42
AUC (optimistic): 0.687 +/- 0.076 (mikro: 0.687) (positive class: Approve)
AUC: 0.628 +/- 0.082 (mikro: 0.628) (positive class: Approve)
AUC (pessimistic): 0.580 +/- 0.088 (mikro: 0.580) (positive class: Approve)
```

Gambar IV.2. Text View Model Confusion Matrix Algoritma C4.5

Perhitungan nilai akurasi data pada algoritma C4.5 sebesar 81,25%, merupakan perhitungan data training terdiri dari 480 record data, 348 data diklasifikasikan Tdk Approve dan 79 data diprediksi Tdk Approve tetapi ternyata Approve, 42 data secara benar diklasifikasikan Approve dan 11 data diprediksi Approve ternyata Tdk Approve.

4.2. Confusion Matrix Algoritma Naïve Bayes

Dari perhitungan terhadap 6 atribut dengan 480 record maka ditemukan 408 record diklasifikasikan Tdk Approve dan 4 record secara benar diklasifikasikan Approve.

Tabel IV.7. Model Tabel Confusion Matrix Algoritma Naïve Bayes

Accuracy: 73,54%

	True Tdk Approve	True Approve	Class Precision
Pred. Tdk Approve	349	117	74,89%
Pred. Approve	10	4	28,57%
Class Recall	97,21%	3,31%	

PerformanceVector

```
PerformanceVector:
accuracy: 73.54% +/- 2.47% (mikro: 73.54%)
ConfusionMatrix:
True: Tdk Approve Approve
Tdk Approve: 349 117
Approve: 10 4
precision: 28.57% (positive class: Approve)
ConfusionMatrix:
True: Tdk Approve Approve
Tdk Approve: 349 117
Approve: 10 4
recall: 3.31% +/- 4.08% (mikro: 3.31%) (positive class: Approve)
ConfusionMatrix:
True: Tdk Approve Approve
Tdk Approve: 349 117
Approve: 10 4
AUC (optimistic): 0.600 +/- 0.056 (mikro: 0.600) (positive class: Approve)
AUC: 0.596 +/- 0.057 (mikro: 0.596) (positive class: Approve)
AUC (pessimistic): 0.592 +/- 0.058 (mikro: 0.592) (positive class: Approve)
```

Gambar IV.3. Text View Model Confusion Matrix Algoritma Naïve Bayes

Perhitungan nilai akurasi data pada algoritma Naïve Bayes sebesar 73,54%, merupakan perhitungan data training terdiri dari 480 record data, 349 data diklasifikasikan Tdk Approve dan 117 data diprediksi Tdk Approve tetapi ternyata Approve, 4 data secara benar diklasifikasikan Approve dan 10 data diprediksi Approve ternyata Tdk Approve.

4.3. Analisa Hasil Komparasi

Dari hasil confusion matrix diatas, selanjutnya dilakukan perhitungan nilai accuracy, precision, dan recall. Perbandingan nilai accuracy, precision, dan recall yang telah dihitung untuk metode C4.5 dan Naïve Bayes dapat dilihat pada tabel berikut

Tabel IV.8. Komparasi Nilai Accuracy, Precision, Recall algoritma C4.5 dan Naïve Bayes

	C4.5	Naïve Bayes
Accuracy		
y	81,25%	73,54%
Precision		
n	81,50%	28,57%
Recall	34,55%	3,33%

Pada tabel diatas terlihat perbandingan nilai accuracy, precision, dan recall dari setiap metode, bahwa nilai secara keseluruhan C4.5 paling tinggi nilainya, dengan nilai accuracy 81,25%, nilai precision 81,50% dan nilai recall 34,55%.

Perbandingan hasil perhitungan nilai AUC untuk metode C4.5 dan Naïve Bayes dapat dilihat pada table berikut.

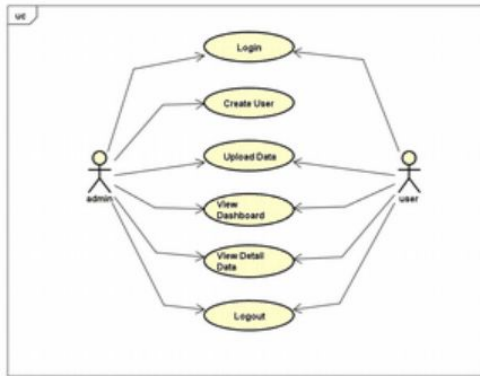
Tabel IV.9. Perbandingan Nilai AUC

AUC	C4.5	Naïve Bayes
	0,687	0,600

Berdasarkan hasil perbandingan akurasi, confusion matrix dan kurva ROC diatas, Algoritma C4.5 yang memiliki nilai yang paling tinggi, sehingga baik digunakan untuk klasifikasi prediksi keberhasilan penawaran kredit bagi para pensiunan PNS.

4.4. Penerapan Algoritma Terpilih

Prototipe yang digunakan dalam penelitian ini dibuat berbasis web dengan menggunakan PHP dan database menggunakan MySQL. Use case diagram untuk prototipe aplikasi sistem adalah sebagai berikut

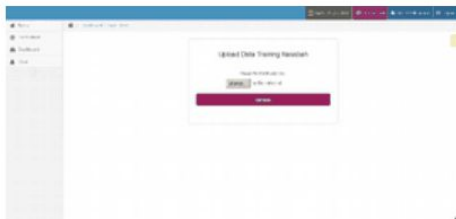


Gambar IV.6. Use Case Diagram Prototipe Aplikasi

Tampilan untuk halaman *Graphical User Interface* (GUI) prototipe prediksi keberhasilan penawaran kredit bagi para pensiunan PNS dapat dilihat pada gambar dibawah.



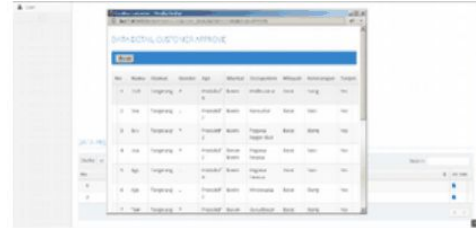
Gambar IV.7. Form Login



Gambar IV.8. Form Upload Data Training Nasabah



Gambar IV.9. Halaman Dashboard



Gambar IV.10. Detail Nasabah Approve

4.5. Pengujian *User Acceptance Test*

Pengujian sistem berbasis *user acceptance testing* yaitu, Pengujian performance prototipe nasabah approve. Tabel berikut merupakan bobot penilaian *mean opinion score* (MOS) yang digunakan.

Tabel IV.10 Mean Opinion Score (MOS)

MOS	Keterangan	Bobot Nilai
SS	Sangat Setuju	5
S	Setuju	4
KS	Kurang Setuju	3
TS	Tidak Setuju	2
TT	Tidak Tahu	1

Hasil penilaian berdasarkan pada masing-masing survey secara subjektif sejumlah 20 responden.

Tabel IV.11 Hasil Jawaban Kuisisioner Kategori *Performance* Prototipe

No	Pertanyaan	SS	S	KS	TS	TT
1.	Prototipe ini mudah digunakan	8	11	1	0	0
2.	Prototipe dapat meload data dengan cepat	9	9	2	0	0
3.	User interface mudah dimengerti	10	3	5	2	0
4.	Tingkat keakurasian data tinggi	6	9	4	1	0
5.	Prototipe sudah memenuhi spesifikasi	7	9	4	0	0
6.	Prototipe sangat membantu dalam memprediksi penawaran kredit	6	10	4	0	0

Dari hasil analisis pengujian *performance* prototipe ditunjukkan pada hasil sebagai berikut :

- 93% responden menyatakan bahwa Prototipe ini mudah digunakan
- 93% responden menyatakan bahwa Prototipe dapat meload data dengan cepat
- 91% responden menyatakan bahwa User interface mudah dimengerti
- 90% responden menyatakan bahwa Tingkat keakurasian data tinggi

5. 91% responden menyatakan bahwa Prototipe sudah memenuhi spesifikasi
6. 91% responden menyatakan bahwa Prototipe sangat membantu dalam menentukan approve nasabah

5. KESIMPULAN

Kesimpulan yang dapat diambil dari penelitian ini adalah berdasarkan hasil evaluasi dan validasi, diketahui bahwa algoritma C4.5 memiliki tingkat akurasi dan performansi yang baik dengan nilai akurasi sebesar 81,25% dan nilai AUC sebesar 0,687. Skor kesimpulan prototipe sistem terhadap pengguna adalah sebagai berikut:

1. 93% responden menyatakan bahwa Prototipe ini mudah digunakan
2. 93% responden menyatakan bahwa Prototipe dapat meload data dengan cepat
3. 91% responden menyatakan bahwa User interface mudah dimengerti
4. 90% responden menyatakan bahwa Tingkat keakurasian data tinggi
5. 91% responden menyatakan bahwa Prototipe sudah memenuhi spesifikasi
6. 91% responden menyatakan bahwa Prototipe sangat membantu dalam menentukan approve nasabah

6. DAFTAR PUSTAKA

- [1] Alpaydm, E., 2012. *Introduction to Machine Learning*. Second., London: The MIT Press.
- [2] Mabur, A., & Lubis, R. (2012). Penerapan *Data mining Untuk Memprediksi Kriteria Nasabah Kredit*. *Jurnal Komputer Dan Informatika (KOMPUTA)*, 1(1), 53–57.
- [3] Ciptohartono, C. C. (2013). Bayes Untuk Menilai Kelayakan Kredit. *Jurnal Universitas Dian Nuswantoro*, 1–6.
- [4] Han, J., & Kamber, M. (2016). *Data Mining: Concepts and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems)*. USA: Elsevier.
- [5] Hardinata, Y. (2013). Analisis Keputusan Pemberian Kredit Modal Kerja Terhadap Usaha Kecil Dan Menengah (Studi Kasus Pada Bank BRI KCP Sukun Malang). *Jurnal Ilmiah Mahasiswa FEB*, 2(1).
- [6] Quinlan, J. R. (1994). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.
- [7] Swarasmaradhana, dkk., 2014. Model Prediksi Tingkat Kelulusan Mahasiswa Dengan Teknik *Data mining* Menggunakan Metode *Decision Tree C4.5*. , 3(2010), hal.1–11.
- [8] Rifai, A. (2016). Kajian Algoritma C4.5, *Naive Bayes*, *Neural Network* dan SVM Dalam Penentuan Kelayakan Pemberian Kredit, (2), 176–182.
- [9] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, NJ: Wiley.
- [10] Lewis, William E. (2009). *Software Testing and Continuous Quality Improvement, Third Edition*. Boca Raton, Florida: CRC Press LLC.
- [11] Perry, William E. (2006). *Effective Methods for Software Testing 3rd Edition*. Indianapolis, Indiana: Wiley Publishing, Inc.