Praktikum 12 - Evaluasi Model Naive Bayes Classifier

December 29, 2021

1 Simulasi Evaluasi Model Naive Bayes Classivier

Pada praktikum kali ini, kita akan mencoba untuk memodifikasi program Naive Bayes Classifier yang telah kita buat sebelumnya agar dapat menghitung performa klasifikasinya. Untuk dapat melakukannya maka diperlukan **training_set** dan **testing_set**. training_set digunakan untuk melatih model Naive Bayes sedangkan testing_set digunakan sebagai bahan ujian model Naive Bayes.

Pada kasus kali ini, model Naive Bayes diminta untuk melakukan analisis sentimen terhadap headline - headline berita terkait masalah keuangan. Dataset yang digunakan bernama **Sentiment Analysis for Financial News** yang dapat Anda akses melalui kaggle. Namun untuk keperluan praktikum kali ini, dataset dimanipulasi sedemikian rupa sehingga dapat digunakan pada praktikum kali ini. Manipulasi meliputi:

1. Jumlah dokumen yang awalnya berjumlah 4837 dikurangi menjadi 100
2. Jumlah kelas yang awalnya ada 3 (positive, neutral, negative) dikurangi menjadi 2 kelas (positive, negative)
3. Komposisi kelas yang awalnya tidak seimbang, pada skenario kali ini dibuat seimbang. 50 dokumen untuk kelas positive dan 50 dokumen untuk kelas negative.
4. Akan dilakukan spliting dataset menjadi training_set dan testing_set dengan rasio 70:30 secara otomatis.

Dataset hasil manipulasi disertakan bersama modul praktikum ini dengan nama dataset.csv.

1.1 Membaca Dataset

Pada tahap ini, data yang ada di dalam dataset.csv dibaca dan disimpan pada variabel dataset. Variabel dataset bertipe List.

```
[1]: #library csv digunakan untuk membaca file csv
     import csv

     #Membuka file dataset.csv. File harus diletakkan di tempat yang sama dengan file␣
      ↪notebook
     file = open('dataset.csv')

     #Membuat object pembaca isi file csv
     csvreader = csv.reader(file)
```

```
#Variabel dataset bertipe list
dataset = []

#Proses pembacaan data pada set dan menyimpannya ke variabel dataset
for baris in csvreader:
        dataset.append(baris)

#Menampilkan isi variabel datasest
for baris in dataset:
    print(baris)
```

['positive', 'In the third quarter of 2010 , net sales increased by 5.2 % to EUR 205.5 mn , and operating profit by 34.9 % to EUR 23.5 mn .']
['positive', 'Operating profit rose to EUR 13.1 mn from EUR 8.7 mn in the corresponding period in 2007 representing 7.7 % of net sales .']
['positive', 'A purchase agreement for 7,200 tons of gasoline with delivery at the Hamina terminal , Finland , was signed with Neste Oil OYj at the average Platts index for this September plus eight US dollars per month .']
['positive', 'Consolidated net sales increased 16 % to reach EUR74 .8 m , while operating profit amounted to EUR0 .9 m compared to a loss of EUR0 .7 m in the prior year period .']
['positive', 'Foundries division reports its sales increased by 9.7 % to EUR 63.1 mn from EUR 57.5 mn in the corresponding period in 2006 , and sales of the Machine Shop division increased by 16.4 % to EUR 41.2 mn from EUR 35.4 mn in the corresponding period in 2006 .']
['positive', "HELSINKI (AFX) - Shares closed higher , led by Nokia after it announced plans to team up with Sanyo to manufacture 3G handsets , and by Nokian Tyres after its fourth-quarter earnings report beat analysts ' expectations , dealers said ."]
['positive', 'Incap Contract Manufacturing Services Pvt Ltd , a subsidiary of Incap Corporation of Finland , plans to double its revenues by 2007-2008 .']
['positive', 'Its board of directors will propose a dividend of EUR0 .12 per share for 2010 , up from the EUR0 .08 per share paid in 2009 .']
['positive', 'Lifetree was founded in 2000 , and its revenues have risen on an average by 40 % with margins in late 30s .']
['positive', '(Filippova) A trilateral agreement on investment in the construction of a technology park in St Petersburg was to have been signed in the course of the forum , Days of the Russian Economy , that opened in Helsinki today .']
['positive', 'Both operating profit and net sales for the six-month period increased , respectively from EUR7 .5 m and EUR655 .5 m , as compared to the corresponding period in 2005 .']
['positive', 'Cargo volume increased by approximately 5 % .']
['positive', 'comparable operating profit totaled EUR 854mn , up from EUR 730mn in 2004 .']
['positive', 'Thereby , the company will expand its offering to North , Central

and South America .']
['positive', "To our members and partners , the use of IT will mostly be apparent in the increased efficiency of the results service , '' observes Perttu Puro from Tradeka ."]
['positive', "The Group 's cash flow from operations will be positive ."]
['positive', 'The mill has long traditions and holds an established position in the markets .']
['positive', 'The move is aimed at boosting sales , cost-efficiency and market share in Finland .']
['positive', 'The parties have also agreed on options that allow UH to extensively implement later on also functionalities of human resources management .']
['positive', "The pilot project proved that RIFD technology is ideal for our purposes '' , comments Olli Saarinen , Material Handling Supervisor at Yara ."]
['positive', 'The planned facility , estimated to cost around $ 814 million , would be the largest biodiesel plant in the world , and use palm oil certified by the Roundtable on Sustainable Palm Oil (RSPO) .']
['positive', 'This combined with foreign investments creates interesting opportunities for Solteq .']
['positive', "This is Done Logistics ' largest order in Norway , the diversified group said ."]
['positive', 'This resulted in improved sales figures in Sweden .']
['positive', 'The borrower was happy to do the roadshow and this paid off as the hit ratio from it was high .']
['positive', 'The company , which makes garden tools , scissors and other consumer goods , said earnings were boosted by 6.9 mln eur of income it received from its 16.7 pct shareholding in Finnish engineering group Wartsila .']
['positive', 'The company anticipates its turnover for the whole 2010 to surpass that of the previous year when it was EUR 67.1 million .']
['positive', 'Atria Group is a leading Scandinavian meat products group .']
['positive', 'Atria will launch a campaign in September to promote the CampoMos brand on the Moscow market , he said .']
['positive', 'Iggesunds Bruk in Sweden is one of the most advanced , fully integrated pulp and paperboard mills in the world .']
['positive', 'In the Homeware area , Fiskars provides leading branded products for home , kitchen , and interior design .']
['positive', 'Jawad Williams scored 23 points in an impressive debut for Jerusalem , with Jason Rich adding 18 and Yuval Naimi chipping in with 16 points for Hapoel , which visits the yellow-and-blue next week .']
['positive', 'More than 200 volunteers from Fiskars , the garden tool company , Home Depot (NYSE : HD) , the city of Baltimore and the neighborhood are working right now to create a garden sanctuary - in a single day .']
['positive', 'NASDAQ-listed Yahoo Inc has introduced a new service that enables Malaysians to take their favorite Internet content and services with them on their mobile phones .']
['positive', "O'Leary 's Material Handling Services , located in Perth , is the leading company in Western Australia that supplies , installs and provides service for tail lifts ."]

['positive', 'One attractive factor about the tablets for media companies is that they provide a new way of making money .']
['positive', "One of Uponor 's promising future products is a hydronic cooling system ."]
['positive', 'Our strong company brands - ESL Shipping , Leipurin , Telko and Kaukomarkkinat - aim to be the market leaders in their sectors .']
['positive', 'By acquiring RCP , we can combine our strong knowhow with the expertise of a solid local company .']
['positive', "EuroChem CFO Andrey Ilyin said : `` This facility marks another important step for EuroChem in securing the long-term financing necessary for our potash and other strategic projects '' ."]
['positive', 'HELSINKI (Thomson Financial) - Kone said it has won four orders in Saudi Arabia , United Arab Emirates and Qatar worth 40 mln eur .']
['positive', '(ADP News) - Finnish handling systems provider Cargotec Oyj (HEL : CGCBV) announced on Friday it won orders worth EUR 10 million (USD 13.2 m) to deliver linkspans to Jordan , Morocco and Ireland .']
['positive', "Merrill Lynch analyst Campbell Morgan upgraded his recommendation on PaperlinX from `` neutral '' to `` buy '' in May ."]
['positive', 'Outotec said it won new orders worth 492.9 mln eur during the quarter , compared with 240.4 mln eur a year earlier .']
['positive', 'President and CEO Mika Vehvil+Ã†inen says the positive signs are first and foremost emerging outside Finland .']
['positive', 'Raute said it has won an order worth around 15 mln eur to supply several plywood production lines to mills operated by Russian wood products company Sveza Group .']
['positive', 'Satama earned Data Management Solutions competency with Business Intelligence specialization recentlyvia the acquisition of Fimentor Oy .']
['positive', 'SysOpen Digia Plc , Press release , 7 February 2006 IBM Finland has rewarded its most distinguished partner companies for 2005 .']
['positive', "The brokerage said 2006 has seen a ` true turning point ' in European steel base prices , with better pricing seen carrying through the second quarter of 2006 ."]
['positive', 'The concept enables a commercially affordable way to manufacture high-quality TCO coated glass for the solar industry .']
['negative', 'Net sales of Finnish Sanoma Learning & Literature , of Finnish media group Sanoma , decreased by 3.6 % in January-June 2009 totalling EUR 162.8 mn , down from EUR 168.8 mn in the corresponding period in 2008 .']
['negative', 'Employing 112 in Finland and 280 abroad , the unit recorded first-quarter 2007 sales of 8.6 mln eur , with an operating loss of 1.6 mln eur .']
['negative', "JP Morgan expects that Scala will lower Nobel Biocare 's growth forecast for 2007 from the current guidance of 23-25 pct , as well as the operating margin target from the current 34-35 pct ."]
['negative', 'Furthermore , sales of new passenger cars and light commercial vehicles in the country declined by 5.4 % year-on-year last month .']
['negative', 'IT services firm TietoEnator was bucking the general trend , holding flat at 22.70 eur , after slipping back from earlier gains .']
['negative', 'Profitability (EBIT %) was 13.6 % , compared to 14.3 % in Q2 2009 .']

['negative', 'Several large stocks tacked lower , however .']
['negative', 'Sales of mid-strength beer decreased by 40 % .']
['negative', "TeliaSonera 's underlying results however included 457 mln skr in positive one-offs , hence the adjusted underlying EBITDA actually amounts to 7.309 bln skr , clearly below expectations , analysts said ."]
['negative', 'Operating profit fell to EUR 35.4 mn from EUR 68.8 mn in 2007 , including vessel sales gain of EUR 12.3 mn .']
['negative', 'Net sales of the Paper segment decreased to EUR 221.6 mn in the second quarter of 2009 from EUR 241.1 mn in the second quarter of 2008 , while operating profit excluding non-recurring items rose to EUR 8.0 mn from EUR 7.6 mn .']
['negative', 'Sales in Finland decreased by 10.5 % in January , while sales outside Finland dropped by 17 % .']
['negative', 'In the Baltic states the company reports net sales of EUR 11.9 mn , down from EUR 14.2 mn , and an operative EBIT of EUR -2.2 mn , down from EUR -1.7 mn .']
['negative', 'Operating profits in the half were 0.8 m , down from 0.9 m as Glisten invested in the brand and the management team .']
['negative', 'Sales in Finland decreased by 2.0 % , and international sales decreased by 9.3 % in terms of euros , and by 15.1 % in terms of local currencies .']
['negative', "The company said that its comparable operating profit for the January-June period fell short of last year 's corresponding performance ."]
['negative', 'Operating result for the 12-month period decreased from the profit of EUR0 .4 m while turnover decreased from EUR5 .6 m , as compared to 2004 .']
['negative', 'HELSINKI Thomson Financial - Shares in Cargotec fell sharply in early afternoon trade after the cargo handling group posted a surprise drop in April-June profits , which overshadowed the large number of new orders received during the three months .']
['negative', 'LONDON MarketWatch -- Share prices ended lower in London Monday as a rebound in bank stocks failed to offset broader weakness for the FTSE 100 .']
['negative', 'The OMX Helsinki index was 0.33 pct lower at 9,364.80 , while the OMX Helsinki CAP portfolio index was down 0.34 pct at 5,029.25 .']
['negative', 'The OMX Helsinki index was down 0.34 pct at 8,256.02 on turnover of 813.191 mln eur .']
['negative', 'Repeats sees 2008 operating profit down y-y (Reporting by Helsinki Newsroom) Keywords : TECNOMEN-RESULTS']
['negative', 'Comparable operating profit totaled EUR 4.7 mn , down from EUR 5.1 mn in the corresponding period in 2005 , representing 7.4 % of net sales .']
['negative', "In Finland 's Hobby Hall 's sales decreased by 10 % , and international sales fell by 19 % ."]
['negative', "Hobby Hall 's sales decrease 26 pct due to implementing a new information system that involved changing in the principal of posting sales ."]
['negative', 'The net sales of the whole fiscal year 2008 will be lower than in 2007 and operating profit is estimated to be negative .']
['negative', 'Sales at the unit slumped last year after the industry was hit by poor snowfall in the major resorts in the winter of 2006-07 .']
['negative', 'The steelmaker said that the drop in profit was explained by the

continuing economic uncertainty , mixed with the current drought in bank lending
, resulting in a decline in demand for its products as customers find it
increasingly difficult to fund operations .']
['negative', "Nokia 's share price fell less than one percent to 18.70 euros ($
25.41) in Helsinki , while Siemens shares fell 1.02 percent to 90.19 euros ($
122.57) in Frankfurt ."]
['negative', 'In February the Elcoteq group sold its St Petersburg facility ;
according to unconfirmed information the reason could have been supply problems
because of the Russian customs service .']
['negative', "`` Operating profit declined mainly due to the increased cost of
wood and recycled fiber and the strengthened euro . ''"]
['negative', 'Operating profit in the fourth quarter went down to EUR3m from
EUR4 .2 m for the corresponding period of 2009 as it included costs of growth
projects .']
['negative', 'According to the company , in addition to normal seasonal
fluctuation the market situation has weakened during autumn 2008 .']
['negative', 'Changes in the market situation and tougher price competition have
substantially reduced demand for bread packaging manufactured at the Kauhava
plant , according to the company .']
['negative', 'In the building and home improvement trade , sales decreased by
6.3 % , totalling EUR 154.1 mn .']
['negative', 'In the building and home improvement trade , net sales totalled
EUR 1,173 mn , down from EUR 1,566 mn a year earlier .']
['negative', 'In the building and home improvement trade , sales decreased by
22.5 % to EUR 201.4 mn .']
['negative', "Finnish Suominen Corporation that makes wipes , nonwovens , and
flexible packaging , expects changes in the market situation to reduce sales of
Suominen 's nonwovens and wet wipes from the previously estimated volumes ."]
['negative', 'ADPnews - Sep 28 , 2009 - Finnish silicon wafers maker Okmetic Oyj
HEL : OKM1V said it will reduce the number of its clerical workers by 22
worldwide as a result of personnel negotiations completed today .']
['negative', '- Demand for fireplace products was lower than expected ,
especially in Germany .']
['negative', 'Currently , the company foresees its pre-tax profit to remain
below the 2009 level when it reached EUR 700,000 USD 934,000 , compared with
previous projections of a slightly better pre-tax profit .']
['negative', 'However , the orders received during the period under review fell
by 17 % quarter-on-quarter from the EUR 213 million recorded in the second
quarter of 2010 .']
['negative', 'The announcement pushed Freenet shares down 6.3 % , or EUR0 .71 ,
in Frankfurt trade to EUR10 .65 as investors gave up hope United Internet AG and
Drillisch would pursue their own takeover and breakup of Freenet .']
['negative', "The terms of the aforementioned funding are considerably below the
Bank 's current CDS levels in the market and have a maturity ranging from 1 to
7.5 years ."]
['negative', 'The total capital of funds managed by the bank decreased by 28 %
to EUR 284mn by the end of September 2008 .']
['negative', 'UPM stock fell 3 percent to EURO 12.49 $ 17.24 in early afternoon

trading in Helsinki .']
['negative', "YIT 's Baltic sales in the first three quarters of 2008 totaled
106.2 million euros , representing a drop of 29 percent year on year ."]
['negative', "Talentum 's net sales in September were smaller than expected ."]
['negative', "The company 's previous estimate was for a lower revenue than the
year-ago EUR 93.9 million ."]
['negative', 'Sales by Seppala diminished by 6 per cent .']

1.2 Mengacak Urutan Dataset

Agar proses pelatihan dan pengujian berjalan secara objektif, maka perlu dilakukan pengacakan
urutan data yang ada pada variabel dataset. Pengacakan dilakukan dengan menggunakan library
"random".

```
[2]:  #Library random digunakan untuk membantu pengacakan data
      import random

      #Proses pengacakan dataset
      random.shuffle(dataset)

      #Menampilkan isi variabel datasest
      for baris in dataset:
          print(baris)
```

['negative', 'LONDON MarketWatch -- Share prices ended lower in London Monday as
a rebound in bank stocks failed to offset broader weakness for the FTSE 100 .']
['positive', 'Satama earned Data Management Solutions competency with Business
Intelligence specialization recentlyvia the acquisition of Fimentor Oy .']
['negative', "Nokia 's share price fell less than one percent to 18.70 euros ($
25.41) in Helsinki , while Siemens shares fell 1.02 percent to 90.19 euros ($
122.57) in Frankfurt ."]
['negative', "YIT 's Baltic sales in the first three quarters of 2008 totaled
106.2 million euros , representing a drop of 29 percent year on year ."]
['negative', 'Furthermore , sales of new passenger cars and light commercial
vehicles in the country declined by 5.4 % year-on-year last month .']
['positive', 'The borrower was happy to do the roadshow and this paid off as the
hit ratio from it was high .']
['positive', 'The concept enables a commercially affordable way to manufacture
high-quality TCO coated glass for the solar industry .']
['positive', "To our members and partners , the use of IT will mostly be
apparent in the increased efficiency of the results service , '' observes Perttu
Puro from Tradeka ."]
['positive', "Merrill Lynch analyst Campbell Morgan upgraded his recommendation
on PaperlinX from `` neutral '' to `` buy '' in May ."]
['positive', 'Outotec said it won new orders worth 492.9 mln eur during the
quarter , compared with 240.4 mln eur a year earlier .']
['positive', 'One attractive factor about the tablets for media companies is

that they provide a new way of making money .']
['negative', 'Employing 112 in Finland and 280 abroad , the unit recorded first-quarter 2007 sales of 8.6 mln eur , with an operating loss of 1.6 mln eur .']
['positive', "EuroChem CFO Andrey Ilyin said : `` This facility marks another important step for EuroChem in securing the long-term financing necessary for our potash and other strategic projects '' ."]
['positive', 'The move is aimed at boosting sales , cost-efficiency and market share in Finland .']
['positive', 'Incap Contract Manufacturing Services Pvt Ltd , a subsidiary of Incap Corporation of Finland , plans to double its revenues by 2007-2008 .']
['positive', 'SysOpen Digia Plc , Press release , 7 February 2006 IBM Finland has rewarded its most distinguished partner companies for 2005 .']
['positive', '(ADP News) - Finnish handling systems provider Cargotec Oyj (HEL : CGCBV) announced on Friday it won orders worth EUR 10 million (USD 13.2 m) to deliver linkspans to Jordan , Morocco and Ireland .']
['positive', 'This combined with foreign investments creates interesting opportunities for Solteq .']
['negative', 'Repeats sees 2008 operating profit down y-y (Reporting by Helsinki Newsroom) Keywords : TECNOMEN-RESULTS']
['positive', 'NASDAQ-listed Yahoo Inc has introduced a new service that enables Malaysians to take their favorite Internet content and services with them on their mobile phones .']
['positive', "The pilot project proved that RIFD technology is ideal for our purposes '' , comments Olli Saarinen , Material Handling Supervisor at Yara ."]
['negative', "JP Morgan expects that Scala will lower Nobel Biocare 's growth forecast for 2007 from the current guidance of 23-25 pct , as well as the operating margin target from the current 34-35 pct ."]
['negative', '- Demand for fireplace products was lower than expected , especially in Germany .']
['negative', 'In the building and home improvement trade , sales decreased by 22.5 % to EUR 201.4 mn .']
['positive', 'The parties have also agreed on options that allow UH to extensively implement later on also functionalities of human resources management .']
['negative', 'HELSINKI Thomson Financial - Shares in Cargotec fell sharply in early afternoon trade after the cargo handling group posted a surprise drop in April-June profits , which overshadowed the large number of new orders received during the three months .']
['negative', 'IT services firm TietoEnator was bucking the general trend , holding flat at 22.70 eur , after slipping back from earlier gains .']
['negative', 'The total capital of funds managed by the bank decreased by 28 % to EUR 284mn by the end of September 2008 .']
['negative', 'However , the orders received during the period under review fell by 17 % quarter-on-quarter from the EUR 213 million recorded in the second quarter of 2010 .']
['negative', 'The OMX Helsinki index was 0.33 pct lower at 9,364.80 , while the OMX Helsinki CAP portfolio index was down 0.34 pct at 5,029.25 .']
['negative', "The company 's previous estimate was for a lower revenue than the

year-ago EUR 93.9 million ."]
['negative', "Finnish Suominen Corporation that makes wipes , nonwovens , and flexible packaging , expects changes in the market situation to reduce sales of Suominen 's nonwovens and wet wipes from the previously estimated volumes ."]
['positive', 'A purchase agreement for 7,200 tons of gasoline with delivery at the Hamina terminal , Finland , was signed with Neste Oil OYj at the average Platts index for this September plus eight US dollars per month .']
['negative', 'Operating profits in the half were 0.8 m , down from 0.9 m as Glisten invested in the brand and the management team .']
['positive', 'HELSINKI (Thomson Financial) - Kone said it has won four orders in Saudi Arabia , United Arab Emirates and Qatar worth 40 mln eur .']
['positive', 'The company , which makes garden tools , scissors and other consumer goods , said earnings were boosted by 6.9 mln eur of income it received from its 16.7 pct shareholding in Finnish engineering group Wartsila .']
['positive', 'Both operating profit and net sales for the six-month period increased , respectively from EUR7 .5 m and EUR655 .5 m , as compared to the corresponding period in 2005 .']
['negative', 'In the Baltic states the company reports net sales of EUR 11.9 mn , down from EUR 14.2 mn , and an operative EBIT of EUR -2.2 mn , down from EUR -1.7 mn .']
['positive', 'Cargo volume increased by approximately 5 % .']
['negative', "Talentum 's net sales in September were smaller than expected ."]
['negative', "TeliaSonera 's underlying results however included 457 mln skr in positive one-offs , hence the adjusted underlying EBITDA actually amounts to 7.309 bln skr , clearly below expectations , analysts said ."]
['negative', 'In February the Elcoteq group sold its St Petersburg facility ; according to unconfirmed information the reason could have been supply problems because of the Russian customs service .']
['positive', 'Lifetree was founded in 2000 , and its revenues have risen on an average by 40 % with margins in late 30s .']
['negative', 'Net sales of the Paper segment decreased to EUR 221.6 mn in the second quarter of 2009 from EUR 241.1 mn in the second quarter of 2008 , while operating profit excluding non-recurring items rose to EUR 8.0 mn from EUR 7.6 mn .']
['positive', "O'Leary 's Material Handling Services , located in Perth , is the leading company in Western Australia that supplies , installs and provides service for tail lifts ."]
['negative', 'UPM stock fell 3 percent to EURO 12.49 $ 17.24 in early afternoon trading in Helsinki .']
['negative', "The terms of the aforementioned funding are considerably below the Bank 's current CDS levels in the market and have a maturity ranging from 1 to 7.5 years ."]
['positive', 'Operating profit rose to EUR 13.1 mn from EUR 8.7 mn in the corresponding period in 2007 representing 7.7 % of net sales .']
['negative', 'Sales by Seppala diminished by 6 per cent .']
['positive', "HELSINKI (AFX) - Shares closed higher , led by Nokia after it announced plans to team up with Sanyo to manufacture 3G handsets , and by Nokian Tyres after its fourth-quarter earnings report beat analysts ' expectations ,

dealers said ."]
['negative', "In Finland 's Hobby Hall 's sales decreased by 10 % , and international sales fell by 19 % ."]
['negative', 'Comparable operating profit totaled EUR 4.7 mn , down from EUR 5.1 mn in the corresponding period in 2005 , representing 7.4 % of net sales .']
['positive', 'Atria Group is a leading Scandinavian meat products group .']
['positive', 'Jawad Williams scored 23 points in an impressive debut for Jerusalem , with Jason Rich adding 18 and Yuval Naimi chipping in with 16 points for Hapoel , which visits the yellow-and-blue next week .']
['negative', 'The OMX Helsinki index was down 0.34 pct at 8,256.02 on turnover of 813.191 mln eur .']
['negative', 'Sales of mid-strength beer decreased by 40 % .']
['negative', 'The announcement pushed Freenet shares down 6.3 % , or EUR0 .71 , in Frankfurt trade to EUR10 .65 as investors gave up hope United Internet AG and Drillisch would pursue their own takeover and breakup of Freenet .']
['negative', 'Net sales of Finnish Sanoma Learning & Literature , of Finnish media group Sanoma , decreased by 3.6 % in January-June 2009 totalling EUR 162.8 mn , down from EUR 168.8 mn in the corresponding period in 2008 .']
['positive', "The brokerage said 2006 has seen a ` true turning point ' in European steel base prices , with better pricing seen carrying through the second quarter of 2006 ."]
['positive', 'Thereby , the company will expand its offering to North , Central and South America .']
['positive', 'Atria will launch a campaign in September to promote the CampoMos brand on the Moscow market , he said .']
['negative', 'The steelmaker said that the drop in profit was explained by the continuing economic uncertainty , mixed with the current drought in bank lending , resulting in a decline in demand for its products as customers find it increasingly difficult to fund operations .']
['negative', "The company said that its comparable operating profit for the January-June period fell short of last year 's corresponding performance ."]
['negative', 'ADPnews - Sep 28 , 2009 - Finnish silicon wafers maker Okmetic Oyj HEL : OKM1V said it will reduce the number of its clerical workers by 22 worldwide as a result of personnel negotiations completed today .']
['negative', 'The net sales of the whole fiscal year 2008 will be lower than in 2007 and operating profit is estimated to be negative .']
['negative', 'Sales in Finland decreased by 2.0 % , and international sales decreased by 9.3 % in terms of euros , and by 15.1 % in terms of local currencies .']
['positive', '(Filippova) A trilateral agreement on investment in the construction of a technology park in St Petersburg was to have been signed in the course of the forum , Days of the Russian Economy , that opened in Helsinki today .']
['positive', 'The company anticipates its turnover for the whole 2010 to surpass that of the previous year when it was EUR 67.1 million .']
['positive', 'In the third quarter of 2010 , net sales increased by 5.2 % to EUR 205.5 mn , and operating profit by 34.9 % to EUR 23.5 mn .']
['negative', "Hobby Hall 's sales decrease 26 pct due to implementing a new

information system that involved changing in the principal of posting sales ."]
['positive', 'More than 200 volunteers from Fiskars , the garden tool company ,
Home Depot (NYSE : HD) , the city of Baltimore and the neighborhood are
working right now to create a garden sanctuary - in a single day .']
['positive', 'Its board of directors will propose a dividend of EUR0 .12 per
share for 2010 , up from the EUR0 .08 per share paid in 2009 .']
['negative', 'Currently , the company foresees its pre-tax profit to remain
below the 2009 level when it reached EUR 700,000 USD 934,000 , compared with
previous projections of a slightly better pre-tax profit .']
['negative', 'Operating profit in the fourth quarter went down to EUR3m from
EUR4 .2 m for the corresponding period of 2009 as it included costs of growth
projects .']
['negative', 'Operating result for the 12-month period decreased from the profit
of EUR0 .4 m while turnover decreased from EUR5 .6 m , as compared to 2004 .']
['positive', "This is Done Logistics ' largest order in Norway , the diversified
group said ."]
['negative', 'Operating profit fell to EUR 35.4 mn from EUR 68.8 mn in 2007 ,
including vessel sales gain of EUR 12.3 mn .']
['positive', 'Foundries division reports its sales increased by 9.7 % to EUR
63.1 mn from EUR 57.5 mn in the corresponding period in 2006 , and sales of the
Machine Shop division increased by 16.4 % to EUR 41.2 mn from EUR 35.4 mn in the
corresponding period in 2006 .']
['positive', 'The mill has long traditions and holds an established position in
the markets .']
['positive', "One of Uponor 's promising future products is a hydronic cooling
system ."]
['positive', 'President and CEO Mika Vehvil+Ãinen says the positive signs are
first and foremost emerging outside Finland .']
['positive', "The Group 's cash flow from operations will be positive ."]
['positive', 'This resulted in improved sales figures in Sweden .']
['positive', 'By acquiring RCP , we can combine our strong knowhow with the
expertise of a solid local company .']
['positive', 'In the Homeware area , Fiskars provides leading branded products
for home , kitchen , and interior design .']
['negative', 'In the building and home improvement trade , sales decreased by
6.3 % , totalling EUR 154.1 mn .']
['negative', 'Profitability (EBIT %) was 13.6 % , compared to 14.3 % in Q2
2009 .']
['positive', 'Iggesunds Bruk in Sweden is one of the most advanced , fully
integrated pulp and paperboard mills in the world .']
['negative', "`` Operating profit declined mainly due to the increased cost of
wood and recycled fiber and the strengthened euro . ''"]
['positive', 'The planned facility , estimated to cost around $ 814 million ,
would be the largest biodiesel plant in the world , and use palm oil certified
by the Roundtable on Sustainable Palm Oil (RSPO) .']
['negative', 'Several large stocks tacked lower , however .']
['positive', 'Raute said it has won an order worth around 15 mln eur to supply
several plywood production lines to mills operated by Russian wood products

company Sveza Group .']
['negative', 'Sales at the unit slumped last year after the industry was hit by poor snowfall in the major resorts in the winter of 2006-07 .']
['positive', 'comparable operating profit totaled EUR 854mn , up from EUR 730mn in 2004 .']
['negative', 'In the building and home improvement trade , net sales totalled EUR 1,173 mn , down from EUR 1,566 mn a year earlier .']
['negative', 'Sales in Finland decreased by 10.5 % in January , while sales outside Finland dropped by 17 % .']
['positive', 'Our strong company brands - ESL Shipping , Leipurin , Telko and Kaukomarkkinat - aim to be the market leaders in their sectors .']
['negative', 'Changes in the market situation and tougher price competition have substantially reduced demand for bread packaging manufactured at the Kauhava plant , according to the company .']
['negative', 'According to the company , in addition to normal seasonal fluctuation the market situation has weakened during autumn 2008 .']
['positive', 'Consolidated net sales increased 16 % to reach EUR74 .8 m , while operating profit amounted to EUR0 .9 m compared to a loss of EUR0 .7 m in the prior year period .']

1.3 Membagi Dataset

Pada tahap ini isi variabel dataset dibagi menjadi dua, yaitu 70% untuk training_set dan 30% untuk testing_set. Besarnya prosesntase sebenarnya tidak ada aturan baku.

```python
#Data pada indeks ke 0 - 69 (70 data) dari dataset digunakan sebagai training_set
training_set = dataset[:70]

#Data pada indeks ke 70 - 99 (30 data) dari dataset digunakan sebagai testing_set
testing_set = dataset[70:]

#Menampilkan isi training_set
print("===TRAINING SET===")
for data in training_set:
    print(data)
print("Banyaknya isi training_set: {}".format(len(training_set)))

print()
#Menampilkan isi testing_set
print("===TESTING SET===")
for data in testing_set:
    print(data)
print("Banyaknya isi testing_set: {}".format(len(testing_set)))
```

[3]:

===TRAINING SET===
['negative', 'LONDON MarketWatch -- Share prices ended lower in London Monday as a rebound in bank stocks failed to offset broader weakness for the FTSE 100 .']

['positive', 'Satama earned Data Management Solutions competency with Business Intelligence specialization recentlyvia the acquisition of Fimentor Oy .']
['negative', "Nokia 's share price fell less than one percent to 18.70 euros ($ 25.41) in Helsinki , while Siemens shares fell 1.02 percent to 90.19 euros ($ 122.57) in Frankfurt ."]
['negative', "YIT 's Baltic sales in the first three quarters of 2008 totaled 106.2 million euros , representing a drop of 29 percent year on year ."]
['negative', 'Furthermore , sales of new passenger cars and light commercial vehicles in the country declined by 5.4 % year-on-year last month .']
['positive', 'The borrower was happy to do the roadshow and this paid off as the hit ratio from it was high .']
['positive', 'The concept enables a commercially affordable way to manufacture high-quality TCO coated glass for the solar industry .']
['positive', "To our members and partners , the use of IT will mostly be apparent in the increased efficiency of the results service , '' observes Perttu Puro from Tradeka ."]
['positive', "Merrill Lynch analyst Campbell Morgan upgraded his recommendation on PaperlinX from `` neutral '' to `` buy '' in May ."]
['positive', 'Outotec said it won new orders worth 492.9 mln eur during the quarter , compared with 240.4 mln eur a year earlier .']
['positive', 'One attractive factor about the tablets for media companies is that they provide a new way of making money .']
['negative', 'Employing 112 in Finland and 280 abroad , the unit recorded first-quarter 2007 sales of 8.6 mln eur , with an operating loss of 1.6 mln eur .']
['positive', "EuroChem CFO Andrey Ilyin said : `` This facility marks another important step for EuroChem in securing the long-term financing necessary for our potash and other strategic projects '' ."]
['positive', 'The move is aimed at boosting sales , cost-efficiency and market share in Finland .']
['positive', 'Incap Contract Manufacturing Services Pvt Ltd , a subsidiary of Incap Corporation of Finland , plans to double its revenues by 2007-2008 .']
['positive', 'SysOpen Digia Plc , Press release , 7 February 2006 IBM Finland has rewarded its most distinguished partner companies for 2005 .']
['positive', '(ADP News) - Finnish handling systems provider Cargotec Oyj (HEL : CGCBV) announced on Friday it won orders worth EUR 10 million (USD 13.2 m) to deliver linkspans to Jordan , Morocco and Ireland .']
['positive', 'This combined with foreign investments creates interesting opportunities for Solteq .']
['negative', 'Repeats sees 2008 operating profit down y-y (Reporting by Helsinki Newsroom) Keywords : TECNOMEN-RESULTS']
['positive', 'NASDAQ-listed Yahoo Inc has introduced a new service that enables Malaysians to take their favorite Internet content and services with them on their mobile phones .']
['positive', "The pilot project proved that RIFD technology is ideal for our purposes '' , comments Olli Saarinen , Material Handling Supervisor at Yara ."]
['negative', "JP Morgan expects that Scala will lower Nobel Biocare 's growth forecast for 2007 from the current guidance of 23-25 pct , as well as the operating margin target from the current 34-35 pct ."]

['negative', '- Demand for fireplace products was lower than expected ,
especially in Germany .']
['negative', 'In the building and home improvement trade , sales decreased by
22.5 % to EUR 201.4 mn .']
['positive', 'The parties have also agreed on options that allow UH to
extensively implement later on also functionalities of human resources
management .']
['negative', 'HELSINKI Thomson Financial - Shares in Cargotec fell sharply in
early afternoon trade after the cargo handling group posted a surprise drop in
April-June profits , which overshadowed the large number of new orders received
during the three months .']
['negative', 'IT services firm TietoEnator was bucking the general trend ,
holding flat at 22.70 eur , after slipping back from earlier gains .']
['negative', 'The total capital of funds managed by the bank decreased by 28 %
to EUR 284mn by the end of September 2008 .']
['negative', 'However , the orders received during the period under review fell
by 17 % quarter-on-quarter from the EUR 213 million recorded in the second
quarter of 2010 .']
['negative', 'The OMX Helsinki index was 0.33 pct lower at 9,364.80 , while the
OMX Helsinki CAP portfolio index was down 0.34 pct at 5,029.25 .']
['negative', "The company 's previous estimate was for a lower revenue than the
year-ago EUR 93.9 million ."]
['negative', "Finnish Suominen Corporation that makes wipes , nonwovens , and
flexible packaging , expects changes in the market situation to reduce sales of
Suominen 's nonwovens and wet wipes from the previously estimated volumes ."]
['positive', 'A purchase agreement for 7,200 tons of gasoline with delivery at
the Hamina terminal , Finland , was signed with Neste Oil OYj at the average
Platts index for this September plus eight US dollars per month .']
['negative', 'Operating profits in the half were 0.8 m , down from 0.9 m as
Glisten invested in the brand and the management team .']
['positive', 'HELSINKI (Thomson Financial) - Kone said it has won four orders
in Saudi Arabia , United Arab Emirates and Qatar worth 40 mln eur .']
['positive', 'The company , which makes garden tools , scissors and other
consumer goods , said earnings were boosted by 6.9 mln eur of income it received
from its 16.7 pct shareholding in Finnish engineering group Wartsila .']
['positive', 'Both operating profit and net sales for the six-month period
increased , respectively from EUR7 .5 m and EUR655 .5 m , as compared to the
corresponding period in 2005 .']
['negative', 'In the Baltic states the company reports net sales of EUR 11.9 mn
, down from EUR 14.2 mn , and an operative EBIT of EUR -2.2 mn , down from EUR
-1.7 mn .']
['positive', 'Cargo volume increased by approximately 5 % .']
['negative', "Talentum 's net sales in September were smaller than expected ."]
['negative', "TeliaSonera 's underlying results however included 457 mln skr in
positive one-offs , hence the adjusted underlying EBITDA actually amounts to
7.309 bln skr , clearly below expectations , analysts said ."]
['negative', 'In February the Elcoteq group sold its St Petersburg facility ;
according to unconfirmed information the reason could have been supply problems

because of the Russian customs service .']
['positive', 'Lifetree was founded in 2000 , and its revenues have risen on an average by 40 % with margins in late 30s .']
['negative', 'Net sales of the Paper segment decreased to EUR 221.6 mn in the second quarter of 2009 from EUR 241.1 mn in the second quarter of 2008 , while operating profit excluding non-recurring items rose to EUR 8.0 mn from EUR 7.6 mn .']
['positive', "O'Leary 's Material Handling Services , located in Perth , is the leading company in Western Australia that supplies , installs and provides service for tail lifts ."]
['negative', 'UPM stock fell 3 percent to EURO 12.49 $ 17.24 in early afternoon trading in Helsinki .']
['negative', "The terms of the aforementioned funding are considerably below the Bank 's current CDS levels in the market and have a maturity ranging from 1 to 7.5 years ."]
['positive', 'Operating profit rose to EUR 13.1 mn from EUR 8.7 mn in the corresponding period in 2007 representing 7.7 % of net sales .']
['negative', 'Sales by Seppala diminished by 6 per cent .']
['positive', "HELSINKI (AFX) - Shares closed higher , led by Nokia after it announced plans to team up with Sanyo to manufacture 3G handsets , and by Nokian Tyres after its fourth-quarter earnings report beat analysts ' expectations , dealers said ."]
['negative', "In Finland 's Hobby Hall 's sales decreased by 10 % , and international sales fell by 19 % ."]
['negative', 'Comparable operating profit totaled EUR 4.7 mn , down from EUR 5.1 mn in the corresponding period in 2005 , representing 7.4 % of net sales .']
['positive', 'Atria Group is a leading Scandinavian meat products group .']
['positive', 'Jawad Williams scored 23 points in an impressive debut for Jerusalem , with Jason Rich adding 18 and Yuval Naimi chipping in with 16 points for Hapoel , which visits the yellow-and-blue next week .']
['negative', 'The OMX Helsinki index was down 0.34 pct at 8,256.02 on turnover of 813.191 mln eur .']
['negative', 'Sales of mid-strength beer decreased by 40 % .']
['negative', 'The announcement pushed Freenet shares down 6.3 % , or EURO .71 , in Frankfurt trade to EUR10 .65 as investors gave up hope United Internet AG and Drillisch would pursue their own takeover and breakup of Freenet .']
['negative', 'Net sales of Finnish Sanoma Learning & Literature , of Finnish media group Sanoma , decreased by 3.6 % in January-June 2009 totalling EUR 162.8 mn , down from EUR 168.8 mn in the corresponding period in 2008 .']
['positive', "The brokerage said 2006 has seen a ` true turning point ' in European steel base prices , with better pricing seen carrying through the second quarter of 2006 ."]
['positive', 'Thereby , the company will expand its offering to North , Central and South America .']
['positive', 'Atria will launch a campaign in September to promote the CampoMos brand on the Moscow market , he said .']
['negative', 'The steelmaker said that the drop in profit was explained by the continuing economic uncertainty , mixed with the current drought in bank lending

, resulting in a decline in demand for its products as customers find it increasingly difficult to fund operations .']
['negative', "The company said that its comparable operating profit for the January-June period fell short of last year 's corresponding performance ."]
['negative', 'ADPnews - Sep 28 , 2009 - Finnish silicon wafers maker Okmetic Oyj HEL : OKM1V said it will reduce the number of its clerical workers by 22 worldwide as a result of personnel negotiations completed today .']
['negative', 'The net sales of the whole fiscal year 2008 will be lower than in 2007 and operating profit is estimated to be negative .']
['negative', 'Sales in Finland decreased by 2.0 % , and international sales decreased by 9.3 % in terms of euros , and by 15.1 % in terms of local currencies .']
['positive', '(Filippova) A trilateral agreement on investment in the construction of a technology park in St Petersburg was to have been signed in the course of the forum , Days of the Russian Economy , that opened in Helsinki today .']
['positive', 'The company anticipates its turnover for the whole 2010 to surpass that of the previous year when it was EUR 67.1 million .']
['positive', 'In the third quarter of 2010 , net sales increased by 5.2 % to EUR 205.5 mn , and operating profit by 34.9 % to EUR 23.5 mn .']
['negative', "Hobby Hall 's sales decrease 26 pct due to implementing a new information system that involved changing in the principal of posting sales ."]
Banyaknya isi training_set: 70

===TESTING SET===
['positive', 'More than 200 volunteers from Fiskars , the garden tool company , Home Depot (NYSE : HD) , the city of Baltimore and the neighborhood are working right now to create a garden sanctuary - in a single day .']
['positive', 'Its board of directors will propose a dividend of EUR0 .12 per share for 2010 , up from the EUR0 .08 per share paid in 2009 .']
['negative', 'Currently , the company foresees its pre-tax profit to remain below the 2009 level when it reached EUR 700,000 USD 934,000 , compared with previous projections of a slightly better pre-tax profit .']
['negative', 'Operating profit in the fourth quarter went down to EUR3m from EUR4 .2 m for the corresponding period of 2009 as it included costs of growth projects .']
['negative', 'Operating result for the 12-month period decreased from the profit of EUR0 .4 m while turnover decreased from EUR5 .6 m , as compared to 2004 .']
['positive', "This is Done Logistics ' largest order in Norway , the diversified group said ."]
['negative', 'Operating profit fell to EUR 35.4 mn from EUR 68.8 mn in 2007 , including vessel sales gain of EUR 12.3 mn .']
['positive', 'Foundries division reports its sales increased by 9.7 % to EUR 63.1 mn from EUR 57.5 mn in the corresponding period in 2006 , and sales of the Machine Shop division increased by 16.4 % to EUR 41.2 mn from EUR 35.4 mn in the corresponding period in 2006 .']
['positive', 'The mill has long traditions and holds an established position in the markets .']

['positive', "One of Uponor 's promising future products is a hydronic cooling system ."]
['positive', 'President and CEO Mika Vehvil+Ã†inen says the positive signs are first and foremost emerging outside Finland .']
['positive', "The Group 's cash flow from operations will be positive ."]
['positive', 'This resulted in improved sales figures in Sweden .']
['positive', 'By acquiring RCP , we can combine our strong knowhow with the expertise of a solid local company .']
['positive', 'In the Homeware area , Fiskars provides leading branded products for home , kitchen , and interior design .']
['negative', 'In the building and home improvement trade , sales decreased by 6.3 % , totalling EUR 154.1 mn .']
['negative', 'Profitability (EBIT %) was 13.6 % , compared to 14.3 % in Q2 2009 .']
['positive', 'Iggesunds Bruk in Sweden is one of the most advanced , fully integrated pulp and paperboard mills in the world .']
['negative', "`` Operating profit declined mainly due to the increased cost of wood and recycled fiber and the strengthened euro . ''"]
['positive', 'The planned facility , estimated to cost around $ 814 million , would be the largest biodiesel plant in the world , and use palm oil certified by the Roundtable on Sustainable Palm Oil (RSPO) .']
['negative', 'Several large stocks tacked lower , however .']
['positive', 'Raute said it has won an order worth around 15 mln eur to supply several plywood production lines to mills operated by Russian wood products company Sveza Group .']
['negative', 'Sales at the unit slumped last year after the industry was hit by poor snowfall in the major resorts in the winter of 2006-07 .']
['positive', 'comparable operating profit totaled EUR 854mn , up from EUR 730mn in 2004 .']
['negative', 'In the building and home improvement trade , net sales totalled EUR 1,173 mn , down from EUR 1,566 mn a year earlier .']
['negative', 'Sales in Finland decreased by 10.5 % in January , while sales outside Finland dropped by 17 % .']
['positive', 'Our strong company brands - ESL Shipping , Leipurin , Telko and Kaukomarkkinat - aim to be the market leaders in their sectors .']
['negative', 'Changes in the market situation and tougher price competition have substantially reduced demand for bread packaging manufactured at the Kauhava plant , according to the company .']
['negative', 'According to the company , in addition to normal seasonal fluctuation the market situation has weakened during autumn 2008 .']
['positive', 'Consolidated net sales increased 16 % to reach EUR74 .8 m , while operating profit amounted to EUR0 .9 m compared to a loss of EUR0 .7 m in the prior year period .']
Banyaknya isi testing_set: 30

1.4 Membuat Fungsi PraPengolahanTeks

Karena tidak dokumen yang digunakan pada sebagai dataset masih dalam bentuk yang tidak normal maka perlu dilakukan normalisasi atau biasa disebut dengan Pra-pengolahan teks. Agar proses pra pengolahan dapat berjalan dengan efisien maka dibuatlah fungsi prapengolahanTeks. Fungsi ini akan menjadikan teks dokumen sebagai input lalu akan melakukan proses normalisasi sehingga menghasilkan teks dokumen yang telah dinormalisasi.

```python
import re
import nltk
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer

def prapengolahanTeks(teks):
    #Fungsi ini digunakan untuk melakukan proses prapengolahan teks dari
 casefolding hingga stemming

    #casefolding
    caseFoldedText = teks.casefold()

    #numbers removal
    polaAngka = r'[\d]'
    noNumberText = re.sub(polaAngka,'',caseFoldedText)

    #punctuation removal
    polaSimbol = r'[^\w\s]'
    noPuncText = re.sub(polaSimbol,'',noNumberText)

    #whitespace removal
    noWhiteSpaceText = noPuncText.strip()

    #word tokenization
    tokenizedText = nltk.word_tokenize(noWhiteSpaceText)

    #stopword removal
    stopWords = set(stopwords.words('english'))
    filteredToken = []
    for token in tokenizedText:
        if token not in stopWords:
            filteredToken.append(token)

    #Lemmanization
    lemma = WordNetLemmatizer()
    lemmatizedToken = []
    for token in filteredToken:
        lemmatizedToken.append(lemma.lemmatize(token))
```

```
    return lemmatizedToken
```

1.5 Membuat dan Melatih Model Naive Bayes

Pada tahap ini dilakukan pembuatan dan pelatihan model Naive Bayes berdasarkan training_set yang sudah dibuat sebelumnya.

```
[5]: #Variabel labelList digunakan untuk menyimpan daftar label
     labelList = []

     #Variabel wordListPositive digunakan untuk menyimpan daftar kata yang masuk
      ↪kelas positive
     wordListPositive = []

     #Variabel wordListNegative digunakan untuk menyimpan daftar kata yang masuk
      ↪kelas negative
     wordListNegative = []

     #bowPositive digunakan untuk menampung Bag of Words kelas positive
     bowPositive = {}

     #bowNegative digunakan untuk menampung Bag of Words kelas negative
     bowNegative = {}

     for dokumen in training_set:
         #Lakukan perulangan untuk setiap dokumen yang ada di training_set

         #Mencatat label ke variabel labelList
         label = dokumen[0]
         labelList.append(label) #Karena label ada pada indeks 0 dokumen

         #Melakukan pra-pengolahan terhadap teks dokumen dengan fungsi
         preProcessedDoc = prapengolahanTeks(dokumen[1]) #Karena teks dokumen pada
     ↪indeks 1 dokumen

         #Pencatatan wordList dan bow
         for token in preProcessedDoc:
             #Lakukan pemeriksaan untuk setiap token yang ada di dalam teks dokumen
             if label == "positive":
                 #Jika label saat ini positive maka simpan token ke wordListPositive
                 wordListPositive.append(token)
                 if token not in bowPositive:
                     #Jika token tidak ada di bowPositive maka simpan token ke
     ↪bowPositive dengan nilai 1
                     bowPositive[token] = 1
                 else:
```

```python
                        #Jika token ada di bowPositive maka lakukan increment 1 nilai
→terhadap nilai token di bowPositive dengan nilai 1
                        bowPositive[token] += 1
            else:
                #Jika label saat ini negative maka simpan token ke wordListNegative
                wordListNegative.append(token)
                if token not in bowNegative:
                        #Jika token tidak ada di bowNegative maka simpan token ke
→bowNegative dengan nilai 1
                        bowNegative[token] = 1
                else:
                        #Jika token ada di bowNegative maka lakukan increment 1 nilai
→terhadap nilai token di bowNegative dengan nilai 1
                        bowNegative[token] += 1


#Menampilkan isi bag of words kelas positive
print("===BAG OF WORDS KELAS POSITIVE===")
for kata in bowPositive:
    print("{}: {}".format(kata,bowPositive[kata]))


print()
#Menampilkan isi bag of words kelas negative
print("===BAG OF WORDS KELAS NEGATIVE===")
for kata in bowNegative:
    print("{}: {}".format(kata,bowNegative[kata]))


print()
#Menghitung jumlah dokumen yang ada di training_set
jmlDokumen = len(training_set)
print("Jumlah dokumen pada training_set: {}".format(jmlDokumen))


print()
#Menghitung jumlah kelas positive di training_set
jmlDokumenPositive = labelList.count("positive")
print("Jumlah dokumen dengan label positive di training_set: {}".
 →format(jmlDokumenPositive))


print()
#Menghitung jumlah kelas negative di training_set
jmlDokumenNegative = labelList.count("negative")
print("Jumlah dokumen dengan label negative di training_set: {}".
 →format(jmlDokumenNegative))


print()
#Menghitung jumlah kata pada kelas positive training_set
jmlKataPositive = len(wordListPositive)
```

```python
print("Jumlah kata yang termasuk label positive di training_set: {}".
 →format(jmlKataPositive))

print()
#Menghitung jumlah kata pada kelas negative training_set
jmlKataNegative = len(wordListNegative)
print("Jumlah kata yang termasuk label negative di training_set: {}".
 →format(jmlKataNegative))

print()
#Menghitung jumlah kata unik yang ada di training_set
semuaKata = []
semuaKata.extend(wordListPositive)
semuaKata.extend(wordListNegative)
kataUnik = set(semuaKata)
jmlKataUnik = len(kataUnik)
print("Jumlah kata unik yang ada di training_set: {}".format(jmlKataUnik))
```

```
===BAG OF WORDS KELAS POSITIVE===
satama: 1
earned: 1
data: 1
management: 2
solution: 1
competency: 1
business: 1
intelligence: 1
specialization: 1
recentlyvia: 1
acquisition: 1
fimentor: 1
oy: 1
borrower: 1
happy: 1
roadshow: 1
paid: 1
hit: 1
ratio: 1
high: 1
concept: 1
enables: 2
commercially: 1
affordable: 1
way: 2
manufacture: 2
highquality: 1
tco: 1
```

```
coated: 1
glass: 1
solar: 1
industry: 1
member: 1
partner: 2
use: 1
mostly: 1
apparent: 1
increased: 4
efficiency: 1
result: 1
service: 6
observes: 1
perttu: 1
puro: 1
tradeka: 1
merrill: 1
lynch: 1
analyst: 2
campbell: 1
morgan: 1
upgraded: 1
recommendation: 1
paperlinx: 1
neutral: 1
buy: 1
may: 1
outotec: 1
said: 7
new: 3
order: 3
worth: 3
mln: 4
eur: 12
quarter: 3
compared: 2
year: 2
earlier: 1
one: 1
attractive: 1
factor: 1
tablet: 1
medium: 1
company: 6
provide: 1
making: 1
money: 1
```

```
eurochem: 2
cfo: 1
andrey: 1
ilyin: 1
facility: 1
mark: 1
another: 1
important: 1
step: 1
securing: 1
longterm: 1
financing: 1
necessary: 1
potash: 1
strategic: 1
project: 2
move: 1
aimed: 1
boosting: 1
sale: 4
costefficiency: 1
market: 2
share: 2
finland: 4
incap: 2
contract: 1
manufacturing: 1
pvt: 1
ltd: 1
subsidiary: 1
corporation: 1
plan: 2
double: 1
revenue: 2
sysopen: 1
digia: 1
plc: 1
press: 1
release: 1
february: 1
ibm: 1
rewarded: 1
distinguished: 1
adp: 1
news: 1
finnish: 2
handling: 3
system: 1
```

```
provider: 1
cargotec: 1
oyj: 2
hel: 1
cgcbv: 1
announced: 2
friday: 1
million: 2
usd: 1
deliver: 1
linkspans: 1
jordan: 1
morocco: 1
ireland: 1
combined: 1
foreign: 1
investment: 2
creates: 1
interesting: 1
opportunity: 1
solteq: 1
nasdaqlisted: 1
yahoo: 1
inc: 1
introduced: 1
malaysian: 1
take: 1
favorite: 1
internet: 1
content: 1
mobile: 1
phone: 1
pilot: 1
proved: 1
rifd: 1
technology: 2
ideal: 1
purpose: 1
comment: 1
olli: 1
saarinen: 1
material: 2
supervisor: 1
yara: 1
party: 1
also: 2
agreed: 1
option: 1
```

```
allow: 1
uh: 1
extensively: 1
implement: 1
later: 1
functionality: 1
human: 1
resource: 1
purchase: 1
agreement: 2
ton: 1
gasoline: 1
delivery: 1
hamina: 1
terminal: 1
signed: 2
neste: 1
oil: 1
average: 2
platts: 1
index: 1
september: 2
plus: 1
eight: 1
u: 1
dollar: 1
per: 1
month: 1
helsinki: 3
thomson: 1
financial: 1
kone: 1
four: 1
saudi: 1
arabia: 1
united: 1
arab: 1
emirate: 1
qatar: 1
make: 1
garden: 1
tool: 1
scissors: 1
consumer: 1
good: 1
earnings: 2
boosted: 1
income: 1
```

```
received: 1
pct: 1
shareholding: 1
engineering: 1
group: 3
wartsila: 1
operating: 3
profit: 3
net: 3
sixmonth: 1
period: 3
respectively: 1
corresponding: 2
cargo: 1
volume: 1
approximately: 1
lifetree: 1
founded: 1
risen: 1
margin: 1
late: 1
oleary: 1
located: 1
perth: 1
leading: 2
western: 1
australia: 1
supply: 1
installs: 1
provides: 1
tail: 1
lift: 1
rose: 1
mn: 4
representing: 1
afx: 1
closed: 1
higher: 1
led: 1
nokia: 1
team: 1
sanyo: 1
g: 1
handset: 1
nokian: 1
tyre: 1
fourthquarter: 1
report: 1
```

```
beat: 1
expectation: 1
dealer: 1
atrium: 2
scandinavian: 1
meat: 1
product: 1
jawad: 1
williams: 1
scored: 1
point: 3
impressive: 1
debut: 1
jerusalem: 1
jason: 1
rich: 1
adding: 1
yuval: 1
naimi: 1
chipping: 1
hapoel: 1
visit: 1
yellowandblue: 1
next: 1
week: 1
brokerage: 1
seen: 2
true: 1
turning: 1
european: 1
steel: 1
base: 1
price: 1
better: 1
pricing: 1
carrying: 1
second: 1
thereby: 1
expand: 1
offering: 1
north: 1
central: 1
south: 1
america: 1
launch: 1
campaign: 1
promote: 1
campomos: 1
```

```
brand: 1
moscow: 1
filippova: 1
trilateral: 1
construction: 1
park: 1
st: 1
petersburg: 1
course: 1
forum: 1
day: 1
russian: 1
economy: 1
opened: 1
today: 1
anticipates: 1
turnover: 1
whole: 1
surpass: 1
previous: 1
third: 1

===BAG OF WORDS KELAS NEGATIVE===
london: 2
marketwatch: 1
share: 5
price: 2
ended: 1
lower: 6
monday: 1
rebound: 1
bank: 4
stock: 2
failed: 1
offset: 1
broader: 1
weakness: 1
ftse: 1
nokia: 1
fell: 7
le: 1
one: 1
percent: 4
euro: 5
helsinki: 7
siemens: 1
frankfurt: 2
yit: 1
```

```
baltic: 2
sale: 19
first: 1
three: 2
quarter: 4
totaled: 2
million: 3
representing: 2
drop: 3
year: 5
furthermore: 1
new: 3
passenger: 1
car: 1
light: 1
commercial: 1
vehicle: 1
country: 1
declined: 1
yearonyear: 1
last: 2
month: 2
employing: 1
finland: 3
abroad: 1
unit: 1
recorded: 2
firstquarter: 1
mln: 4
eur: 22
operating: 8
loss: 1
repeat: 1
see: 1
profit: 8
yy: 1
reporting: 1
newsroom: 1
keywords: 1
tecnomenresults: 1
jp: 1
morgan: 1
expects: 2
scala: 1
nobel: 1
biocare: 1
growth: 1
forecast: 1
```

```
current: 4
guidance: 1
pct: 6
well: 1
margin: 1
target: 1
demand: 2
fireplace: 1
product: 2
expected: 2
especially: 1
germany: 1
building: 1
home: 1
improvement: 1
trade: 3
decreased: 8
mn: 14
thomson: 1
financial: 1
cargotec: 1
sharply: 1
early: 2
afternoon: 2
cargo: 1
handling: 1
group: 3
posted: 1
surprise: 1
apriljune: 1
overshadowed: 1
large: 1
number: 2
order: 2
received: 2
service: 2
firm: 1
tietoenator: 1
bucking: 1
general: 1
trend: 1
holding: 1
flat: 1
slipping: 1
back: 1
earlier: 1
gain: 1
total: 1
```

```
capital: 1
fund: 2
managed: 1
end: 1
september: 2
however: 2
period: 4
review: 1
quarteronquarter: 1
second: 3
omx: 3
index: 3
cap: 1
portfolio: 1
company: 3
previous: 1
estimate: 1
revenue: 1
yearago: 1
finnish: 4
suominen: 2
corporation: 1
make: 1
wipe: 2
nonwovens: 2
flexible: 1
packaging: 1
change: 1
market: 2
situation: 1
reduce: 2
wet: 1
previously: 1
estimated: 2
volume: 1
half: 1
glisten: 1
invested: 1
brand: 1
management: 1
team: 1
state: 1
report: 1
net: 6
operative: 1
ebit: 1
talentum: 1
smaller: 1
```

```
teliasonera: 1
underlying: 2
result: 2
included: 1
skr: 2
positive: 1
oneoffs: 1
hence: 1
adjusted: 1
ebitda: 1
actually: 1
amount: 1
bln: 1
clearly: 1
expectation: 1
analyst: 1
said: 4
february: 1
elcoteq: 1
sold: 1
st: 1
petersburg: 1
facility: 1
according: 1
unconfirmed: 1
information: 2
reason: 1
could: 1
supply: 1
problem: 1
russian: 1
custom: 1
paper: 1
segment: 1
excluding: 1
nonrecurring: 1
item: 1
rose: 1
upm: 1
trading: 1
term: 3
aforementioned: 1
funding: 1
considerably: 1
cd: 1
level: 1
maturity: 1
ranging: 1
```

```
seppala: 1
diminished: 1
per: 1
cent: 1
hobby: 2
hall: 2
international: 2
comparable: 2
corresponding: 3
turnover: 1
midstrength: 1
beer: 1
announcement: 1
pushed: 1
freenet: 2
investor: 1
gave: 1
hope: 1
united: 1
internet: 1
ag: 1
drillisch: 1
would: 1
pursue: 1
takeover: 1
breakup: 1
sanoma: 2
learning: 1
literature: 1
medium: 1
januaryjune: 2
totalling: 1
steelmaker: 1
explained: 1
continuing: 1
economic: 1
uncertainty: 1
mixed: 1
drought: 1
lending: 1
resulting: 1
decline: 1
customer: 1
find: 1
increasingly: 1
difficult: 1
operation: 1
short: 1
```

```
performance: 1
adpnews: 1
sep: 1
silicon: 1
wafer: 1
maker: 1
okmetic: 1
oyj: 1
hel: 1
okmv: 1
clerical: 1
worker: 1
worldwide: 1
personnel: 1
negotiation: 1
completed: 1
today: 1
whole: 1
fiscal: 1
negative: 1
local: 1
currency: 1
decrease: 1
due: 1
implementing: 1
system: 1
involved: 1
changing: 1
principal: 1
posting: 1

Jumlah dokumen pada training_set: 70

Jumlah dokumen dengan label positive di training_set: 33

Jumlah dokumen dengan label negative di training_set: 37

Jumlah kata yang termasuk label positive di training_set: 435

Jumlah kata yang termasuk label negative di training_set: 496

Jumlah kata unik yang ada di training_set: 561
```

1.6 Melakukan Pengujian Terhadap Model

Pada tahap ini, model yang telah dibuat dan dilatih akan diuji dengan testing_set.

```python
[6]: #Variabel output_final akan menyimpan item berupa list, dengan indeks 0 sebagai
     ↪label_asli dan indeks 1 sebagai label_tebakan
     output_final = []

     for dokumen in testing_set:
         #Melakukan iterasi untuk setiap dokumen yang ada di testing_set

         #Mendapatkan label asli dari setiap dokumen yang ada di testing_set
         label_asli = dokumen[0] #karena label ada pada indeks 0 di dokumen testing_set

         #Mendapatkan teks dokumen yang ada di testing_set
         dokumen_testing = dokumen[1] #karena teks dokumen ada di indeks 1 dokumen
     ↪testing_set

         #Melakukan pra-pengolahan terhadap dokumen_testing
         preProcessed_dokumen_testing = prapengolahanTeks(dokumen_testing)

         #Mendefinisikan probabilitas suatu dokumen masuk ke suatu kelas
         probabilitasPositive = jmlDokumenPositive / jmlDokumen
         probabilitasNegative = jmlDokumenNegative / jmlDokumen

         #Menghitung nilai probabilitas masing-masing token
         for token in preProcessed_dokumen_testing:
             #Lakukan iterasi untuk setiap token yang ada di
     ↪preProcessed_dokumen_testing

             #Menghitung nilai pembilang untuk token di masing-masing kelas
             pembilangPositive = bowPositive.get(token,0) + 1
             pembilangNegative = bowNegative.get(token,0) + 1

             #Menghitung nilai penyebut untuk token di masing-masing kelas
             penyebutPositive = jmlKataPositive + jmlKataUnik
             penyebutNegative = jmlKataNegative + jmlKataUnik

             #Menghitung nilai pembilang / penyebut untuk setiap kelas
             rasioPositive = pembilangPositive / penyebutPositive
             rasioNegative = pembilangNegative / penyebutNegative

             #Menghitung probabilitas total untuk masing-masing kelas
             probabilitasPositive = probabilitasPositive * rasioPositive
             probabilitasNegative = probabilitasNegative * rasioNegative

         #Membandingkan nilai probabilitas untuk masing - masing kelas
         if probabilitasPositive > probabilitasNegative:
             #Jika probabilitas kelas positive lebih besar daripada kelas negative
     ↪maka nilai tebakan = "positive"
             label_tebakan = "positive"
```

```python
    elif probabilitasPositive < probabilitasNegative:
        #Jika probabilitas kelas negative lebih besar daripada kelas positive
↪maka nilai tebakan = "negative"
        label_tebakan = "negative"
    else:
        #Jika nilai probabilitasPositive sama dengan probabilitasNegative maka
↪nilai tebakan = -
        label_tebakan = "-"

    #Menyimpan label_asli dan label_tebakan ke dalam output_temporary
    output_temporary = []
    output_temporary.append(label_asli)
    output_temporary.append(label_tebakan)

    #Menyimpan output_temporary ke dalam output_final
    output_final.append(output_temporary)

#Menampilkan label_asli dan label_tebakan untuk setiap dokumen yang ada di
↪testing_set
print("---------------------------------------")
print("|No.\t| Label Asli\t| Label Tebakan\t|")
print("---------------------------------------")
for indeks, tebakan in enumerate(output_final):
    #Lakukan iterasi untuk setiap item yang ada di output_final
    print("|{}\t| {}\t| {}\t|".format(indeks+1,tebakan[0],tebakan[1]))
print("---------------------------------------")
```

```
---------------------------------------
|No.     | Label Asli    | Label Tebakan |
---------------------------------------
|1       | positive      | positive      |
|2       | positive      | negative      |
|3       | negative      | positive      |
|4       | negative      | negative      |
|5       | negative      | negative      |
|6       | positive      | positive      |
|7       | negative      | negative      |
|8       | positive      | negative      |
|9       | positive      | positive      |
|10      | positive      | negative      |
|11      | positive      | negative      |
|12      | positive      | negative      |
|13      | positive      | negative      |
|14      | positive      | positive      |
|15      | positive      | positive      |
|16      | negative      | negative      |
|17      | negative      | positive      |
```

```
|18       | positive    | positive      |
|19       | negative    | negative      |
|20       | positive    | positive      |
|21       | negative    | negative      |
|22       | positive    | positive      |
|23       | negative    | negative      |
|24       | positive    | negative      |
|25       | negative    | negative      |
|26       | negative    | negative      |
|27       | positive    | positive      |
|28       | negative    | negative      |
|29       | negative    | negative      |
|30       | positive    | negative      |
----------------------------------------
```

1.7   Menghitung Performa Model Naive Bayes

Pada bagian ini, performa model Naive Bayes untuk melakukan klasifikasi sentimen. Performa dihitung dengan mengetahui nilai True Positive, True Negative, False Positive, dan False Negative. Setelah didapatkan semuanya maka akan dihitung akurasinya, presisi, recall, dan f_measure-nya.

```python
[7]: #Variabel true_positive menyimpan jumlah kejadian label_asli dan label_tebakan␣
     ↪sama-sama bernilai positive
     true_positive = output_final.count(["positive","positive"])

     #Variabel true_negative menyimpan jumlah kejadian label_asli dan label_tebakan␣
     ↪sama-sama bernilai negative
     true_negative = output_final.count(["negative","negative"])

     #Variabel false_positive menyimpan jumlah kejadian label_asli bernilai negative␣
     ↪sedangkan label_tebakan bernilai positive
     false_positive = output_final.count(["negative","positive"])

     #Variabel false_negative menyimpan jumlah kejadian label_asli bernilai positive␣
     ↪sedangkan label_tebakan bernilai negative
     false_negative = output_final.count(["positive","negative"])

     print("True Positive\t: {}".format(true_positive))
     print("True Negative\t: {}".format(true_negative))
     print("False Positive\t: {}".format(false_positive))
     print("False Negative\t: {}".format(false_negative))

     print()
     #Variabel akurasi menyimpan hasil perhitungan akurasi
     akurasi = (true_positive + true_negative) / (true_positive + true_negative +␣
     ↪false_positive + false_negative)
```

```python
#Variabel presisi menyimpan hasil perhitungan presisi
presisi = true_positive / (true_positive + false_positive)

#Variabel recall menyimpan hasil perhitungan recall
recall = true_positive / (true_positive + false_negative)

#Variabel f_measure menyimpan hasil perhitungan f_measure
f_measure = 2 * (presisi * recall) / (presisi + recall)

print("Akurasi\t\t: %.2f persen" % (akurasi*100))
print("Presisi\t\t: %.2f persen" % (presisi*100))
print("Recall\t\t: %.2f persen" % (recall*100))
print("F-1 Mesure\t: %.2f persen" % (f_measure*100))
```

```
True Positive    : 9
True Negative    : 11
False Positive   : 2
False Negative   : 8

Akurasi          : 66.67 persen
Presisi          : 81.82 persen
Recall           : 52.94 persen
F-1 Mesure       : 64.29 persen
```

[ ]: