# Disclaimer Detection

Basically all different Disclaimer sections can be divided into two categories:

First page disclaimer: disclaimer reference or analyst certification on first page.

End of document disclosure section usually contains Analyst Certification, Rating Definition/Distribution, Disclaimer and sometimes contains Valuation and Risks, Historical Performance charts

## Examples:

1. Please see page(s) 2 - 3 for important disclosures and analyst certification. 1
2. **Important Disclosures**

This report constitutes a compendium report (covers six or more subject companies). As such, Brean Capital, LLC. chooses to provide specific disclosures for the companies mentioned by reference. To access current disclosures for the all companies in this report, clients should refer to Disclosure Site or contact your Brean Capital, LLC. representative for additional information. Trading (212)-702-6527 Member FINRA, SIPC www.breancapital.com

……..

Brean Capital, LLC Stock Rating System

**Buy - Expected to appreciate by at least 10% within the next 12 months.**

**Hold - Fully valued, not expected to appreciate or decline materially within the next 12 months.**

**Sell - Expected to decline by at least 10% within the next 12 months.**

**IB Serv./ Past 12Mos. Rating Category Count Percent Count Percent BUY 143 70.10% 17 11.89% HOLD 56 27.45% 2 3.57% SELL 5 2.45% 0 0.00% NOT RATED Note: Stock price** volatility may cause temporary non-alignment of some ratings with some target prices.

**Analyst Certification**

I, Lucas Pipes, hereby certify that the views expressed in this research report accurately reflect our personal views about any and all of the subject securities or issuers referred to in this document. The analyst and associate analyst further certify that they have not received and will not be receiving direct or indirect compensation in exchange for expressing the recommendation contained in this publication.

Disclaimers

Brean Capital, LLC. Equity Research 2

1. Page 21

Befimmo

Recommendation history for BEFIMMO

Date Recommendation Target price Price at change date 16-Dec-13 Hold 54.00 49.31 02-Aug-13 Hold 52.88 50.00 17-May-13 Hold 51.92 51.87 15-Mar-13 Hold 50.00 45.48 27-Jun-12 Hold 48.08 42.84 17-Feb-12 Hold 51.92 47.65 18-Nov-11 Hold 54.81 50.23 21-Sep-11 Hold 56.73 55.18 24-Aug-11 Hold 58.65 53.99

Source: Factset & ESN, price data adjusted for stock splits. This chart shows Bank Degroof continuing coverage of this stock; the current analyst may or may not have covered it over the entire period. Current analyst: Jean-Marie Caucheteux (since 23/11/2009)

**53 52 51 50 49 48 47 46 45 44 Dec 12 Jan 13 Feb 13 Mar 13 Apr 13 May 13 Jun 13 Jul 13 Aug 13 Sep 13 Oct 13 Nov 13 Dec 13 Jan 14**

Price history Target price history

Buy Accumulat Hold Reduce Sell Not rated

# Algorithm details

## Training set format

Each training file follow the rules:

* \f – is used as page delimiter
* \r\n is paragraph deilimiter, each line is a single paragraph
* Every disclaimer paragraph starts with “<DISCLAIMER>” tag

Logistic regression algorithm is used to predict class of paragraph.

Noisy paragraphs (disclaimer paragraphs that were seen before) are excluded from classification.

## Pargraph Features

1 |w 10 11 12 19 28 31 47 50 75 87 106 119 184 289 349 961 2073 2200 |f 1:5.0 2:2.0 3:0.4

1 |w 2 3 4 5 7 9 10 15 16 17 19 22 25 27 29 30 31 33 36 37 39 41 42 46 47 50 51 55 61 62 65 66 69 71 73 78 84 86 90 97 103 108 119 126 128 130 134 140 143 144 147 156 169 180 181 185 187 199 201 248 264 277 292 324 338 344 354 359 374 451 469 503 511 531 573 605 627 748 762 777 781 788 809 822 852 868 875 895 915 924 936 967 1029 1041 1082 1204 1222 1246 1292 1347 1403 2316 2538 2831 |f 1:7.0 2:6.0 3:0.857142857142857

-1 |w 47 423 1483 |f 1:3.0 2:0.0 3:0

1 |w 50 |f 1:6.0 2:5.0 3:0.833333333333333

1. Namespace **|w**: unigrams.

Every word is normalized:

* Token is converted to special token class if it is being matched by regex

|  |  |
| --- | --- |
| Token Class | Regex |
| \_\_number\_\_ | "^(CAD|C|US|EUR|HK)?[#$()+\\-,./\\\\]\*\\d+([#$()+\\-,./\\\\]+\\d+)\*[#$()+\\-,./\\\\XY]\*$" |
| \_\_date\_\_ | "^("  + "\\d{1,2}/\\d{1,2}/\\d{4}"  + "|(\\dQ|Q\\d)?[F|C]Y\\d+[EA]?"  + "|\\dQ(\\d+[EA]?)?"  + "|(\\d+[\\-/])?(JAN(UARY)?|FEB(RUARY)?|MAR(CH)?|APR(IL)?|MAY|JUN(E)?|JUL(Y)?|AUG(UST)?|SEP(TEMBER)?|NOV(EMBER)?|DEC(EMBER)?)([\\-/]\\d+)?"  + ")$" |
| \_\_alphanumerical\_\_ | "^[A-Z]\*\\d+([A-Z]+\\d+)\*[A-Z]\*$ |
| \_\_email\_\_ | "^[A-Z0-9.\_%+-]+@[A-Z0-9.-]+\\.[A-Z]+$" |
| \_\_url\_\_ | "^(https?://|www\\.)|\\.(com|org|net|edu|gov)$" |
| \_\_delimiter\_\_ | "^[^A-Z\\d]+$" |

* All non-alphabet characters are removed and all characters are converted to lower case for the rest of the tokens

Currently I use top 3000 tokens ranked using Chi Square Statistics. Could be changed to any features selection algorithm

1. Namespace **|f:** page features.

2.1 Page count

2.2 Zero based page index

2.3 {Zero based page index} / {Page count}

## Model

Logistic regression with L2 regularization

> vw.exe -d disclaimer.vw.train -f disclaimer.vw --loss\_function logistic --passes 25 -c -l2

## Output

|  |
| --- |
| final\_regressor = disclaimer.vw  Num weight bits = 18  learning rate = 2  initial\_t = 0  power\_t = 0.5  decay\_learning\_rate = 1  using cache\_file = disclaimer.vw.train.cache  ignoring text input in favor of cache input  num sources = 1  average since example example current current current  loss last counter weight label predict features  0.859488 0.859488 3 3.0 -1.0000 -0.0543 7  0.762933 0.666379 6 6.0 -1.0000 -0.2389 5  0.544203 0.281726 11 11.0 -1.0000 -1.0627 4  0.616968 0.689733 22 22.0 -1.0000 -2.2566 5  0.662976 0.708985 44 44.0 1.0000 -4.1706 46  0.555053 0.444621 87 87.0 -1.0000 -1.5199 6  0.474314 0.393575 174 174.0 1.0000 -0.1539 5  0.401195 0.328077 348 348.0 -1.0000 -1.9341 5  0.316585 0.231974 696 696.0 -1.0000 -4.5250 13  0.281711 0.246837 1392 1392.0 -1.0000 -0.8258 4  0.245560 0.209410 2784 2784.0 -1.0000 -5.7898 13  0.205108 0.164656 5568 5568.0 -1.0000 -2.9986 5  0.170958 0.136802 11135 11135.0 1.0000 4.7800 13  0.146171 0.121382 22269 22269.0 -1.0000 -1.7765 7  0.123808 0.101444 44537 44537.0 -1.0000 -7.4563 5  0.106994 0.090180 89073 89073.0 -1.0000 -5.4782 2  0.094796 0.082598 178146 178146.0 -1.0000 -10.6674 10  0.085112 0.075427 356291 356291.0 -1.0000 -5.9153 3  0.076906 0.068700 712582 712582.0 -1.0000 -8.6627 7  0.071117 0.065327 1425163 1425163.0 -1.0000 -8.1390 14  0.066847 0.062577 2850326 2850326.0 -1.0000 -3.8906 6  0.063785 0.060723 5700651 5700651.0 -1.0000 -7.1576 2  0.061607 0.059430 11401301 11401301.0 -1.0000 -0.8008 7  finished run  number of examples = 11957675  weighted example sum = 1.19577e+007  weighted label sum = -7.83758e+006  average loss = 0.0614939  best constant = -0.655443  total feature number = 151672525 |

## Test

> vw.exe -d disclaimer.vw.train -t -i ./disclaimer.vw -p disclaimer.pred

|  |
| --- |
| only testing  Num weight bits = 18  learning rate = 10  initial\_t = 1  power\_t = 0.5  predictions = disclaimer.pred  using no cache  Reading datafile = disclaimer.vw.train  num sources = 1  average since example example current current current  loss last counter weight label predict features  175.844114 175.844114 3 3.0 -1.0000 -0.8710 7  93.814023 11.783931 6 6.0 -1.0000 -4.6023 5  100.840839 109.273018 11 11.0 -1.0000 -5.6951 4  200.518419 300.195999 22 22.0 -1.0000 -2.4550 5  146.652010 92.785601 44 44.0 1.0000 26.0825 46  148.901040 151.202374 87 87.0 -1.0000 -12.7606 6  160.112408 171.323776 174 174.0 1.0000 6.9369 5  130.853610 101.594812 348 348.0 -1.0000 -4.8427 5  137.381221 143.908832 696 696.0 -1.0000 -6.4521 13  134.642624 131.904026 1392 1392.0 -1.0000 -0.3051 4  135.653497 136.664369 2784 2784.0 -1.0000 -10.1172 13  133.328636 131.003776 5568 5568.0 -1.0000 -5.6512 5  131.839319 130.349734 11135 11135.0 1.0000 11.5138 13  135.441642 139.044288 22269 22269.0 -1.0000 -1.7634 7  137.813044 140.184553 44537 44537.0 -1.0000 -11.7865 5  139.666685 141.520367 89073 89073.0 -1.0000 -7.5617 2  139.994445 140.322205 178146 178146.0 -1.0000 -13.2336 10  139.948835 139.903225 356291 356291.0 -1.0000 -7.8795 3  finished run  number of examples = 478307  weighted example sum = 478307  weighted label sum = -313503  average loss = 140.186  best constant = -0.655447  total feature number = 6066901 |

## Performace of the model on training set

tp = 74615, fp = 1951, tn = 393954, fn = 7787

precision: 97.45%, recall: 90.55%, fscore: 93.87%

## Detection of Disclaimer Pages

The next step after classification is to predict disclaimer pages.

Page is considered to be disclaimer page if one of the following conditions is satisfied:

* Length of predicted disclaimer is more than half of the page length
* Previous page is disclaimer page and the next page is disclaimer page
* Previous page is disclaimer page and current page is the last page

For every disclaimer page:

* All paragraphs from first disclaimer paragraph to the end of the page are considered to be disclaimer paragraphs.
* And if previous page was disclaimer page then all paragraphs are considered to be disclaimer.

## Algorithm Performance and Cutoff value

Computed on 507 unseen files, training set 2000 files

**3000 tokens, Cutoff 0.0**

tp = 5662360, fp = 198261, fn = 68799, tn = 13147848

precision = 96.62, recall = 98.80, fscore = 97.70

**3000 tokens, Cutoff 1.0**

tp = 5654571, fp = 125048, fn = 76588, tn = 13221061

precision = 97.84, recall = 98.66, fscore = 98.25

**3000 tokens, Cutoff 2.0**

tp = 5645753, fp = 82302, fn = 85406, tn = 13263807

precision = 98.56, recall = 98.51, fscore = 98.54

**3000 tokens, Cutoff 3.0**

tp = 5639386, fp = 70009, fn = 91773, tn = 13276100

precision = 98.77, recall = 98.4, fscore = 98.59

**3000 tokens, Cutoff 4.0**

tp = 5627902, fp = 36852, fn = 103257, tn = 13309257

precision = 99.35, recall = 98.2, fscore = 98.77

**3000 tokens, Cutoff 5.0**

tp = 5611360, fp = 32000, fn = 119799, tn = 13314109

precision = 99.43, recall = 97.91, fscore = 98.67

## Tools

DisclaimerTrainingSetToVW – is used to convert training set to vowpal wabbit training file

DisclaimerTestPerformance – is used to test performance on training set

DisclaimerAnnotate – is used to annotate documents