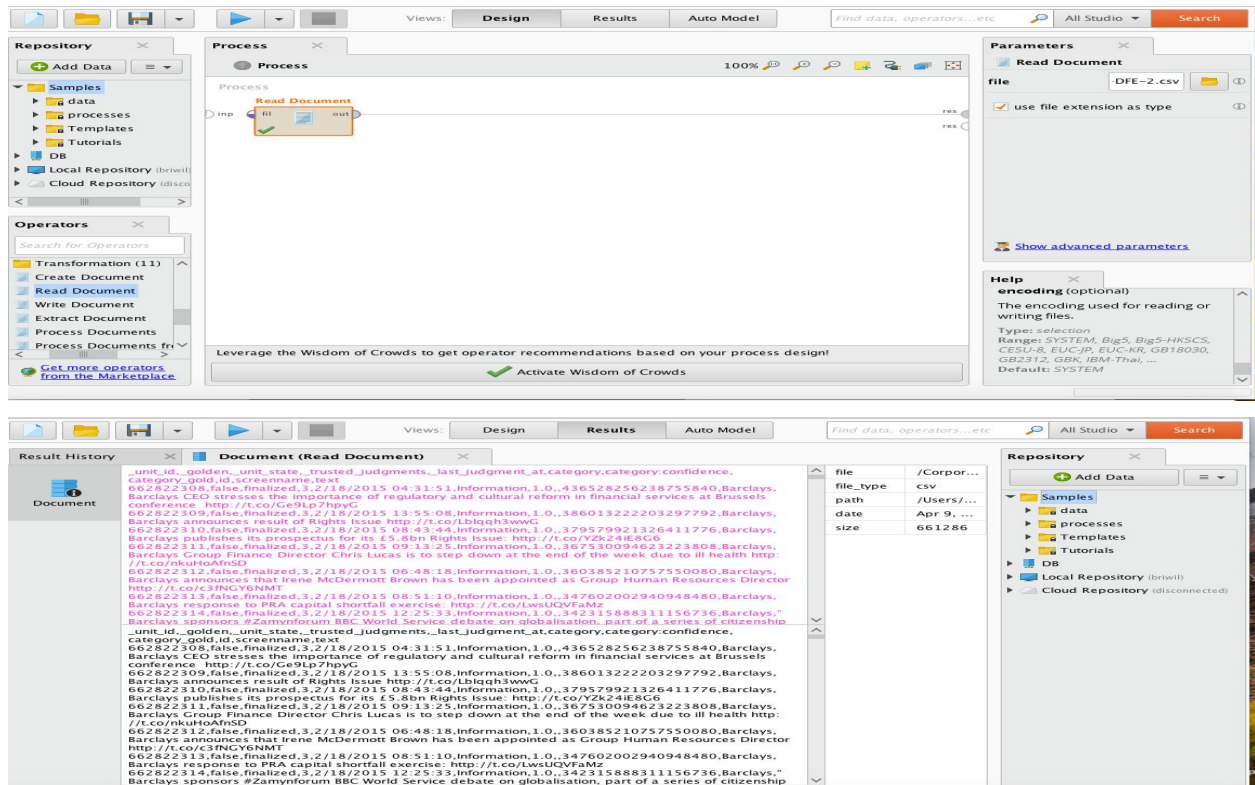
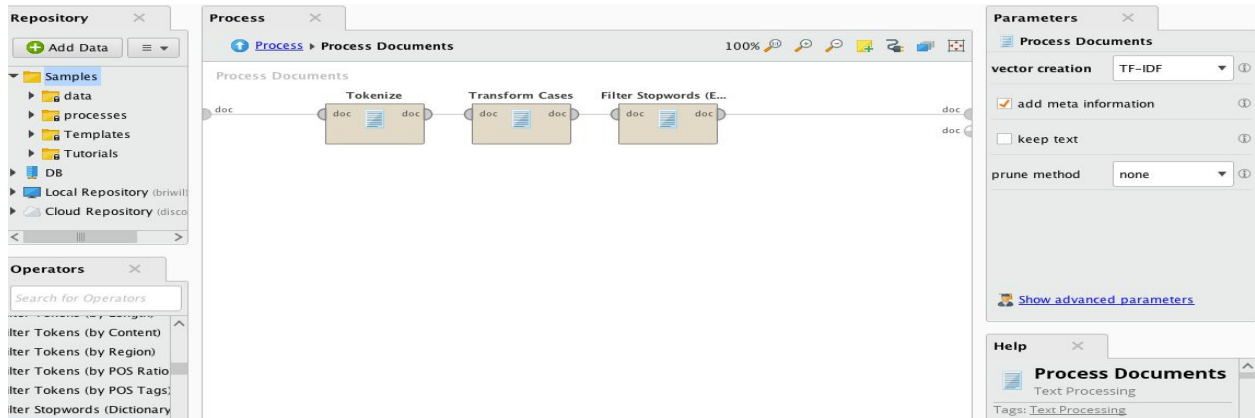


After connecting the Read Document out node to the res node, I then selected the Corporate-messaging-DFE-2.csv file. From here I connected the Read Document operator to the output to import the text, and return to the design window.

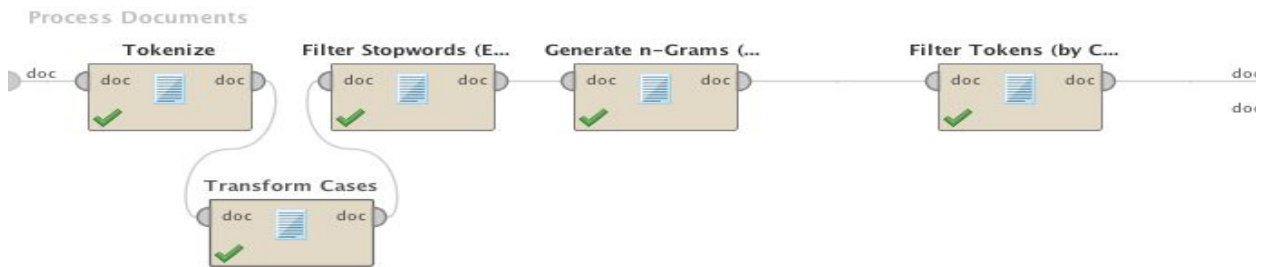


With the Read Document, the Process Documents, Transform Cases, Filter Stopwords(English), and the Tokenize operator connected to the doc node I ran the process, which displays Word, Attribute Name, Total Occurrences & Document Occurrences as they appear in the corporate-messaging-DFE-2 document.



Word	Attribu...	Total O...	Docum...
aa	aa	1	1
aaa	aaa	1	1
aacr	aacr	2	1
aahrpp	aahrpp	6	1
aampayuc	aampayuc	1	1
aaron	aaron	5	1
aay	aay	1	1
aazdrovl	aazdrovl	1	1
ab	ab	1	1
abando...	abando...	2	1
abbott	abbott	1	1
aberron	aberron	1	1
abetterny	abetterny	1	1
ability	ability	3	1
able	able	1	1

I added the Generate n-Grams (Terms) operator to the sub process within the Process Documents operator. I added Filter Tokens (by Content) and set the parameters to conditions = contains and string = \_ (underscore). We can see the name Aaron Gurwitz appears five times in one document.



Word	Attribu...	Total O...	Docum...
aa_false	aa_false	1	1
aaa_pro...	aaa_pro...	1	1
aacr_look	aacr_look	2	1
aahrpp...	aahrpp...	4	1
aahrpp...	aahrpp...	2	1
aampay...	aampay...	1	1
aaron_g...	aaron_g...	5	1
aay_o	aay_o	1	1
aazdrov...	aazdrov...	1	1
ab_false	ab_false	1	1

I created a new process and then created a new document and entered the tweet text file into the Parameter window. I selected the Documents to Data operator, and set text attribute to = text. I the setup Aylien by going to Connections -> Manage Connections. Add a connection for Aylien. Then I added the Analyze Sentiment operator.

The screenshot shows the RapidMiner Studio interface. The central 'Process' window displays a workflow: 'Create Document' (out) -> 'Documents to Data' (doc, doc) -> 'Analyze Sentiment' (Eka, Eka). The 'Parameters' window for the 'Process' operator is open, showing various settings: logverbosity (init), logfile, resultfile, random seed (2001), send mail (never), and encoding (SYSTEM). The 'Help' window for 'Process' is also visible, stating it is the root operator of every process.

ExampleSet (1 example, 4 special attributes, 1 regular attribute) Filter (1 / 1 examples): all

how the data in a table

Row No.	polarity_co...	subjectivit...	polarity	subjectivity	text
1	0.983	0.574	neutral	subjective	Barclays CE...

I then imported the corporate-messaging-DFE-2 file and connected it to a Set Role operator, and a Sample operator.

The screenshot shows the RapidMiner Studio interface. The central 'Process' window displays a workflow: 'Retrieve Corporate...' (out) -> 'Set Role' (exa, ori) -> 'Sample' (exa, ori). The 'Parameters' window for the 'Set Role' operator is open, showing settings: attribute name (screenname), target role (label), and set additional roles (Edit List (0)...). The 'Help' window for 'Set Role' is also visible, listing tags like Label, Target, Id, Class, etc.

Result History

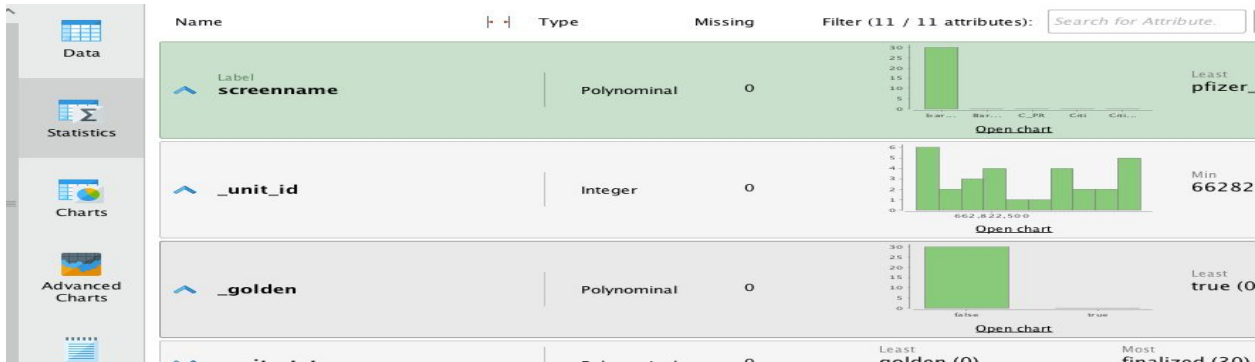
ExampleSet (Sample)

ExampleSet (30 examples, 1 special attribute, 10 regular attributes) Filter (30 / 30 examples): all

Row No.	screenname	_unit_id	_golden	_unit_state	_trusted_ju...	_last_judg...	category
1	barclayswea...	662822409	false	finalized	3	Feb 18, 20...	Action
2	barclayswea...	662822412	false	finalized	3	Feb 18, 20...	Informatic
3	barclayswea...	662822413	false	finalized	3	Feb 18, 20...	Action
4	barclayswea...	662822416	false	finalized	3	Feb 18, 20...	Action
5	barclayswea...	662822421	false	finalized	3	Feb 18, 20...	Informatic
6	barclayswea...	662822434	false	finalized	3	Feb 18, 20...	Informatic
7	barclayswea...	662822450	false	finalized	3	Feb 18, 20...	Informatic
8	barclayswea...	662822451	false	finalized	3	Feb 18, 20...	Informatic
9	barclayswea...	662822468	false	finalized	3	Feb 18, 20...	Informatic
10	barclayswea...	662822473	false	finalized	3	Feb 18, 20...	Informatic
11	barclayswea...	662822485	false	finalized	3	Feb 18, 20...	Informatic
12	barclayswea...	662822491	false	finalized	3	Feb 18, 20...	Informatic
13	barclayswea...	662822493	false	finalized	3	Feb 18, 20...	Informatic
14	barclayswea...	662822502	false	finalized	3	Feb 18, 20...	Informatic

DB

- Local Repository (briwil)
  - data (briwil)
    - Corporate-messaging-DFE
    - Corporate-messaging-DFE
  - processes (briwil)
    - assignment 1 (briwil - v1, 4/14/1...
    - Corporate-messaging-DFE-2
    - Newbie (briwil - v1, 3/14/18 10:5...
  - Cloud Repository (disconnected)



After connecting to the Analyze Sentiment operator I ran the process. I then made a Histogram Color chart with polarity. You can see that Merck had the most positive posts, though nearly ties with Citi in negative posts.

Result History

ExampleSet (Sample)

ExampleSet (249 examples, 5 special attributes, 10 regular attributes) Filter (249 / 249 examples): all

Row No.	screenname	polarity_co...	subjectivit...	polarity	subjectivity	_unit_id	_golden
1	Barclays	0.983	0.904	neutral	objective	662822308	false
2	Barclays	0.974	0.984	neutral	subjective	662822309	false
3	Barclays	0.952	0.921	neutral	subjective	662822310	false
4	Barclays	0.900	1.000	neutral	objective	662822311	false
5	Barclays	0.959	0.626	neutral	subjective	662822312	false
6	Barclays	0.957	0.556	neutral	objective	662822313	false
7	Barclays	0.925	1.000	neutral	objective	662822314	false
8	Barclays	0.960	1.000	neutral	subjective	662822315	false
9	Barclays	0.973	1.000	neutral	subjective	662822316	false
10	barclayswea...	0.551	1.000	negative	objective	662822409	false
11	barclayswea...	0.970	1.000	neutral	objective	662822412	false
12	barclayswea...	0.940	1	neutral	subjective	662822413	false
13	barclayswea...	0.726	1	neutral	subjective	662822416	false
14	barclayswea...	0.508	1	negative	subjective	662822421	false

