

Portfolio_1

閲覧履歴から自分の好む見出しを
機械学習でピックアップ

概要

・訓練用データ

4月16～25日までのまとめサイトのRSS情報を取得 →ラベル0(不正解ラベル)

3月1日～4月26日のまとめサイト閲覧履歴を取得 →ラベル1(正解ラベル)

テストデータ

4月27日のまとめサイトRSS情報と閲覧履歴

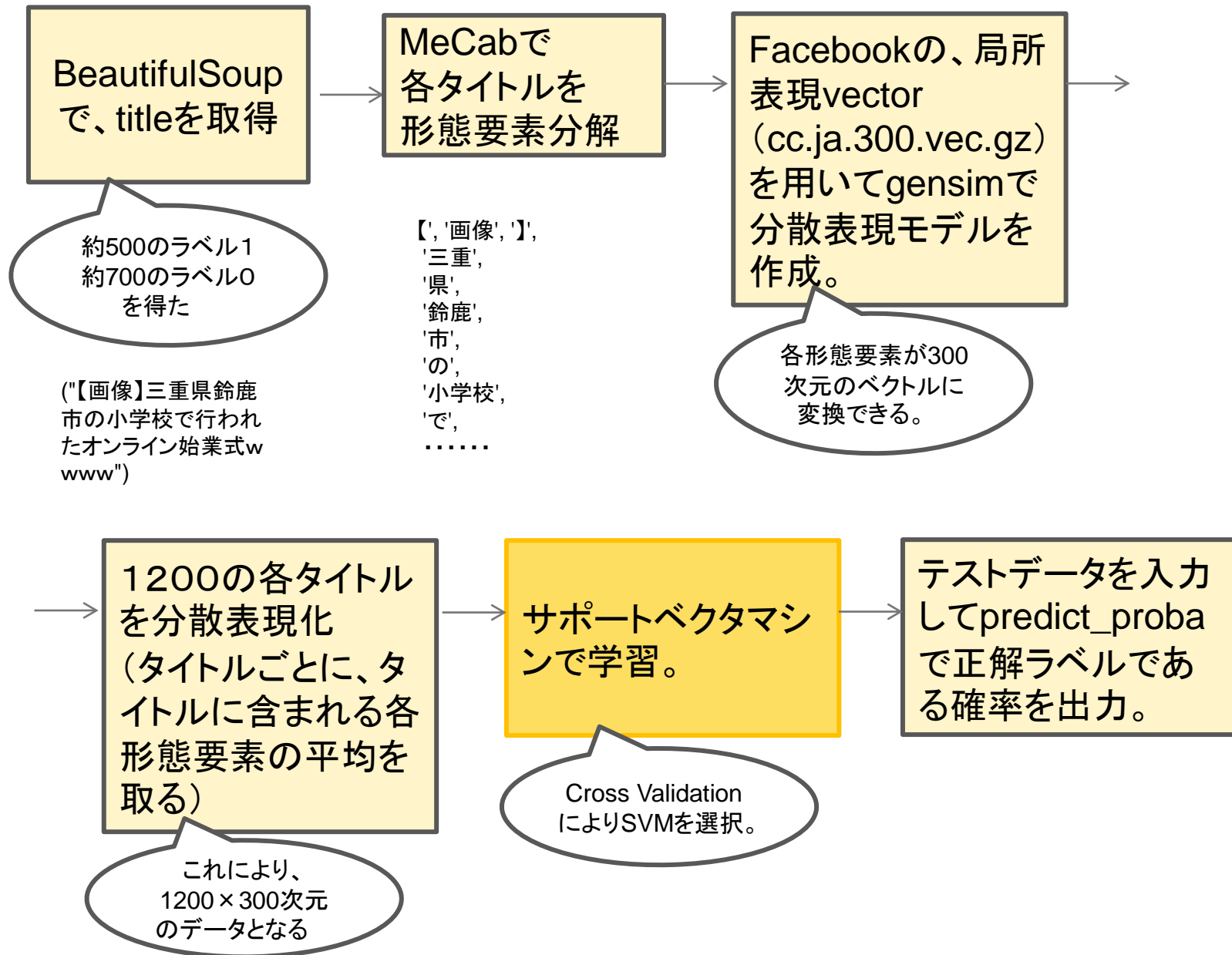
トレインデータの一部。

・Titleに関しては最低限の前処理を行っている(サイト名の排除など)。

・urlをIDとして活用し、重複排除等の処理をしている。

	title	url	date	target
0	山中教授「マスク無しでジョギングするのはウイルスを撒き散らしている可能性がある」	http://blog.livedoor.jp/nanjstu/arc	Mon, 20 Apr	0
1	「背番号5」←真っ先に誰を思い浮かべた？	http://blog.livedoor.jp/nanjstu/arc	Mon, 20 Apr	0
2	ADHDワイ「あと5分で家出なアカン……せや！」	http://himasoku.com/archives/52	Fri, 24 Apr	0
3	【一覧流出】国際機関、ほぼ中国人に乗っ取られていたw	http://crx7601.com/archives/566	Tue, 21 Apr	0
4	【速報】政府支給マスクの虫・毛・カビ 全部バヨのデマだった？菅官房長官が説明！！	http://crx7601.com/archives/566	Fri, 24 Apr	0
5	ガチの「アメリカ料理」ってなんかあるの？	http://burusoku-vip.com/archives	Tue, 21 Apr	0
6	東京から岩手に引越して即焼死した男性、コロナ警戒したマンション住人に拒まれ、役所からも転入届を断られてた	http://blog.livedoor.jp/dqnplus/arc	Mon, 20 Apr	0
7	【10万円支給】閣議決定やり直し 安倍首相「1日も早く国民の手元に届けられるよう全力で取り組む」	http://military38.com/archives/54	Mon, 20 Apr	0
8	入籍したから歴代の元カノに結婚報告した結果www ww	http://burusoku-vip.com/archives	Tue, 21 Apr	0
9	ドラッグストア店員なんだけど客にカードを返そうとしたら、「移ったらどうすんのよ汚い…」と言われた	http://oniyomech.livedoor.biz/arch	Mon, 20 Apr	0
10	【悲報】ワイのアホパッパ、10万円は受け取らないと言い出す……	http://inazumanews2.com/archive	Fri, 24 Apr	0
11	国民「安倍無能、辞めろ」謎の勢力「代わりは？野党に任せるのか？」	http://military38.com/archives/54	Tue, 21 Apr	0
12	大阪府民「コロナなのにレストランが営業してる！通報しなきゃ！！」→休業へ	http://burusoku-vip.com/archives	Tue, 21 Apr	0
13	何を言ってるかわからないと思うが、米国ではコロナのおかげで18年ぶりに銃乱射のない3月だったそうだ	http://himasoku.com/archives/52	Sat, 18 Apr	0

機械学習のフロー



テストデータを入力したところ、まずまずの結果が得られた。

不正解ラベルの可能性が低いサンプル・・・軍事、歴史、コロナ、経済関連

不正解ラベルの可能性が高いサンプル・・・生活関連が多い

ヨッメの呼び方で夫婦仲分かる説wwwwwwwwwwww

頭良くなる方法教えてくれwwwwwwwwwwww

【急募】深夜徘徊の楽しみ方教えてくれ

ワイ「10万……」ボソッ 父「金いらないから働いてないんだろ？」

【画像】買ってきたネギの保管方法がこちらwwwwww

【愕然】寿司屋で3500円頼んだ結果……

太るために筋トレしてるんやが食事辛いわ…

【悲報】ワイ筋トレ民、80kgでベンチプレスしてたら隣のデブが100kgでやり出すwwwwww

上司「ワイ君ゲーマーなんだって？PS4持ってるなら一緒にやろうよw」ワイ「あはは…」

【画像】弁当によく付いているアレがガチャになってしまった件wwwwww

【悲報】職場の女先輩にどうぶつの森一緒にやろうと誘われてしまうwwwwwwww

友達「トイレ借りていい？」ワイ「ええで^^でも勝手に流さないでね？」

【悲報】無職男(20)、特殊詐欺の受け子に挑戦もスーツが似合わなすぎで即バレ…身分証は食べて隠す

0.806663

0.8003

0.79830

0.7987

0.788

0.784308

0.781709

0.77792

0.776728

0.774293

0.773789

0.772087

気の迷いでクリックしそうな傾向の見出し群。

日本軍「沖縄戦線に大和投入するンゴ…！！」

韓国軍兵士が軍事機密暗号を「カカオトーク」に漏洩し懲戒処分に…ネットは「今の軍は烏合の衆」！

日本の感染数がガチで収束傾向にある件

【新型コロナ】緊急事態宣言 政府内「来月6日全面解除は困難」

米ボーイング社、ブラジルの航空機大手エンブラエルとの事業統合を中止！

【速報】中国で甲殻類に感染する謎のウイルス蔓延開始wwwwww

地球を侵略してきた宇宙人←こいつら

【新型コロナ】インターハイ史上初の中止

強襲揚陸艦「アメリカ」の飛行甲板では海兵隊員らがゴルフ練習で休息！

イエメン停戦を1カ月延長、サウジアラビア主導のアラブ連合軍…新型コロナ確認後も衝突続く！

日銀が追加金融緩和 上限なく国債購入 政府の経済対策に連動

アメリカ太平洋空軍初のF-35A部隊、アラスカに誕生…第354戦闘航空団！

0.316691

0.313848

0.30180

0.296

0.2967

0.2865

0.261841

0.213593

0.209896

0.189485

0.138579

黄色のマーカーは閲覧履歴に残っていたリンク先。

数値は、不正解ラベルの可能性を示す