

Appendix A.

```
{ "id": "10103259157", "score": 1, "content": "东西没看到，说是烂了又退回了，最后苦苦追了大半个月把钱退回来了", "referenceName": "张裕 (CHANGYU) 红酒 优选级干红葡萄酒 750ml", "isMobile": true, "userLevelName": "钻石会员", "userLevelId": "105", "referenceTime": "2017-01-15 21:25:02", "days": 19, "replyCount": 1, "usefulVoteCount": 1, "uselessVoteCount": 0, "userClient": 4, "userClientShow": "来自京东Android客户端", "creationTime": "2017-02-03 19:04:52" }
```

Figure A.1: One online review sample from JD.com.

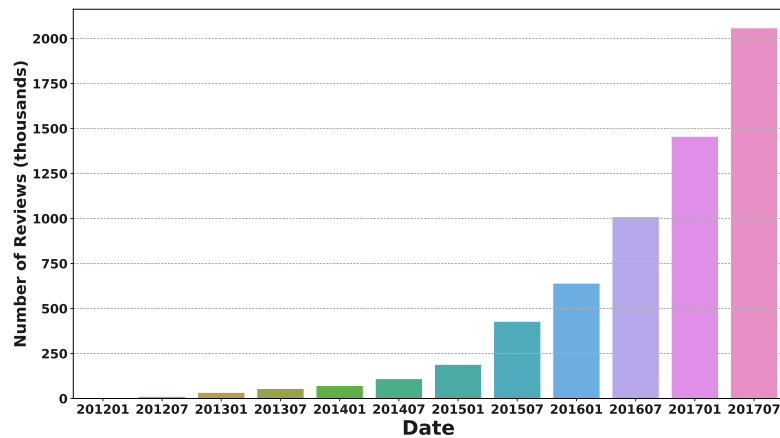


Figure A.2: The number of all firms' product reviews in half-year granularity. Around 2014, given the credit to ubiquity of mobile Internet industry, there is an exponential growth in online product reviews with the rapid rise of online shopping activities.

Table A.1: Review Firms

Sectors	Codes	Short Names	#Product	#Review
Home Appliance	002677SZ	ZHEJIANG MEIDA	74	2352
Home Appliance	600261SH	ZHEJIANG YANKON	165	2779
Home Appliance	600983SH	Whirlpool	248	2884
Home Appliance	002681SZ	FENDA	28	17035
Home Appliance	002076SZ	CNLIGHT	427	20118
Home Appliance	000921SZ	HISENSE HA	213	43347
Home Appliance	000533SZ	SHUNNA	930	48576
Home Appliance	002668SZ	HOMA	69	55924
Home Appliance	002508SZ	ROBAM APPLIANCES	964	90841
Home Appliance	000418SZ	LITTLESWAN	1225	91222
Home Appliance	002543SZ	VANWARD ELECTRIC	473	93509
Home Appliance	600336SH	AUCMA	583	113729
Home Appliance	600060SH	HXDQ	1525	128656
Home Appliance	000521SZ	CHML	859	131582
Home Appliance	600839SH	CHANGHONG	1923	152903
Home Appliance	000541SZ	FSL	649	154792
Home Appliance	000016SZ	KONKA GROUP	1784	204636
Home Appliance	002035SZ	VATTI	1387	246855
Home Appliance	000651SZ	GREE	1593	310044
Home Appliance	002403SZ	ASD	2297	405346
Home Appliance	600890SH	CRED HOLDING	5998	431188
Home Appliance	000100SZ	TCL	5997	579173
Home Appliance	002242SZ	JOYOUNG	2677	673990
Home Appliance	000333SZ	MIDEA GROUP	5980	1580060
Home Appliance	002032SZ	SUPOR	5991	1664441
Garment	002687SZ	GIUSEPPE	239	1395
Garment	002569SZ	BUSEN CORP.	171	1513
Garment	600107SH	mailyard	684	1663
Garment	002574SZ	MING JEWELRY	453	2148

Table A.1 continued from previous page

Sectors	Codes	Short Names	#Product	#Review
Garment	002731SZ	CUIHUA JEWELRY	407	3239
Garment	002044SZ	MEINIAN ONEHEALTH	596	5077
Garment	002291SZ	SATURDAY	1480	5113
Garment	600398SH	HLA	5997	6811
Garment	002036SZ	LIANCHUANG ELECTRON	36	7934
Garment	600177SH	Youngor	3712	10602
Garment	002612SZ	LANCY	1104	12465
Garment	601718SH	Jihua Group	652	15264
Garment	002003SZ	WEIXING	91	17575
Garment	002269SZ	METERSBONWE	3213	25368
Garment	002345SZ	CHJ	1882	29763
Garment	002397SZ	MENDALE	795	31116
Garment	601566SH	JOEONE	4517	58651
Garment	002563SZ	SEMIR	5060	74330
Garment	600884SH	NBSS	3794	76070
Garment	603555SH	Guirenniao	6030	142494
Garment	300005SZ	TOREAD	6000	207577
Garment	002327SZ	FUANNA	1113	215161
Garment	002293SZ	LUOLAI	5825	243486
Garment	603001SH	AoKang	5956	244251
Garment	600400SH	Hongdou Industrial	3974	293600
Garment	600137SH	langshagufen	5996	1663330
Garment	002029SZ	SEPTWOLVES	5997	1969384
Food	002387SZ	VISIONOX	19	2832
Food	000848SZ	CHENGDELOLO	17	4110
Food	002481SZ	SHUANGTA FOOD	25	7390
Food	002702SZ	HAIXIN FOODS	44	17560
Food	002726SZ	LONGDA	25	21537
Food	002330SZ	DELISI	138	22501
Food	600419SH	TIAN RUN	335	27267
Food	002515SZ	JINZI HAM	173	32537

Table A.1 continued from previous page

Sectors	Codes	Short Names	#Product	#Review
Food	600300SH	VVFB	95	33546
Food	002661SZ	KEMEN	101	40311
Food	002507SZ	FULING ZHACAI	96	44595
Food	002732SZ	YANTANG DAIRY	66	47143
Food	600305SH	Hengshun Vinegar	184	47699
Food	600597SH	BRIGHT DAIRY	238	48116
Food	000639SZ	XIWANG	79	48205
Food	002216SZ	SANQUAN FOODS	333	51833
Food	000716SZ	NANFANG BLACK SESAME	81	54021
Food	600429SH	SANYUAN	496	70381
Food	600073SH	SMAC	430	76033
Food	600298SH	ANGEL	322	105062
Food	002695SZ	HUANGSHANGHUANG	400	117454
Food	002570SZ	BEINGMATE	3369	143891
Food	002582SZ	HAOXIANGNI	414	182331
Food	000895SZ	SHUANGHUI	1085	185673
Food	002557SZ	CHACHA FOOD CO.,LTD	402	191174
Food	603288SH	HAI TIAN	159	194193
Food	600887SH	YILI	4454	488522
Food	300146SZ	BY-HEALTH	3384	734398
Food	600737SH	Cofco Tunhe	6000	981302
Wine	603369SH	King's Luck	348	2143
Wine	600702SH	SHE DE	57	2779
Wine	002461SZ	ZHUJIANG BREWERY	16	3476
Wine	600199SH	AGSW	68	5029
Wine	600238SH	HAINAN YEDAO	51	11538
Wine	600543SH	GSMG	169	16934
Wine	600197SH	YLT	173	17063
Wine	601579SH	KUALJI	729	18894
Wine	600779SH	SCSF	396	24856
Wine	000729SZ	YANJING BREWERY	103	25121

Table A.1 continued from previous page

Sectors	Codes	Short Names	#Product	#Review
Wine	000799SZ	JGJC	682	27919
Wine	600559SH	lao bai gan jiu	311	28326
Wine	600600SH	TSINGTAO BREWERY	139	40665
Wine	000596SZ	GUJING DISTILLERY	570	50731
Wine	600059SH	GYLS	1210	53641
Wine	000869SZ	CHANGYU	723	80161
Wine	600809SH	Shanxi Fen Wine	1786	136564
Wine	600519SH	KWEICHOW MOUTAI	5555	137011
Wine	002304SZ	YANGHE	1686	158968
Wine	000568SZ	LUZHOU LAO JIAO	3131	245394
Wine	000858SZ	WULIANGYE	4015	288223

Table A.9: Tendency Features

Feature	Definition
$W_n Tendency_t^{posW}$	$\sum_{i=1}^N ReviewPos_{ti}$
$W_n Tendency_t^{posW} Diff_t$	$W_n Tendency_t^{posW} - W_n Tendency_{t-1}^{posW}$
$W_n Tendency_t^{posW} DiffRatio_t$	$\frac{W_n Tendency_t^{posW} - W_n Tendency_{t-1}^{posW}}{W_n Tendency_{t-1}^{posW}}$
$W_n Tendency_t^{posW} Average_t$	$\frac{W_n Tendency_t^{posW}}{W_n Review_t}$
$W_n Tendency_t^{posW} AverageDiff_t$	$W_n Tendency_t^{posW} Average_t - W_n Tendency_{t-1}^{posW} Average_{t-1}$
$W_n Tendency_t^{posW} DiffH_t^m$	$W_n Tendency_t^{posW} - W_m Tendency_{t-n}^{posW}$
$W_n Tendency_t^{posW} AverageDiffH_t^m$	$W_n Tendency_t^{posW} Average_t - W_m Tendency_{t-n}^{posW} Average_{t-n}$
$W_n Tendency_t^{negW}$	$\sum_{i=0}^N ReviewNeg_{ti}$
$W_n Tendency_t^{negW} Diff_t$	$W_n Tendency_t^{negW} - W_n Tendency_{t-1}^{negW}$
$W_n Tendency_t^{negW} DiffRatio_t$	$\frac{W_n Tendency_t^{negW} - W_n Tendency_{t-1}^{negW}}{W_n Tendency_{t-1}^{negW}}$
$W_n Tendency_t^{negW} Average_t$	$\frac{W_n Tendency_t^{negW}}{W_n Review_t}$
$W_n Tendency_t^{negW} AverageDiff_t$	$W_n Tendency_t^{negW} Ratio_t - W_n Tendency_{t-1}^{negW} Ratio_{t-1}$
$W_n Tendency_t^{negW} DiffH_t^m$	$W_n Tendency_t^{negW} - W_m Tendency_{t-n}^{negW}$
$W_n Tendency_t^{negW} AverageDiffH_t^m$	$W_n Tendency_t^{negW} Average_t - W_m Tendency_{t-n}^{negW} Average_{t-n}$

Table A.9 continued from previous page

Feature	Definition
$W_n Tendency_t^{word}$	$\frac{W_n Tendency_t^{pos} - W_n Tendency_t^{neg}}{W_n Tendency_t^{pos} + W_n Tendency_t^{neg}}$
$W_n Tendency_t^{word} Diff_t$	$W_n Tendency_t^{word} - W_n Tendency_{t-1}^{word}$
$W_n Tendency_t^{word} DiffH_t^m$	$W_n Tendency_t^{word} - W_n Tendency_{t-n}^{word}$
$W_n Tendency_t^{posR}$	$\sum_{i=0}^N \{1 \text{ if } ReviewPost_i > ReviewNeg_i \text{ else } 0\}$
$W_n Tendency_t^{posR} Diff_t$	$W_n Tendency_t^{posR} - W_n Tendency_{t-1}^{posR}$
$W_n Tendency_t^{posR} DiffRatio_t$	$\frac{W_n Tendency_t^{posR} - W_n Tendency_{t-1}^{posR}}{W_n Tendency_{t-1}^{posR}}$
$W_n Tendency_t^{posR} Ratio_t$	$\frac{W_n Tendency_t^{posR}}{W_n Review_t}$
$W_n Tendency_t^{posR} RatioDiff_t$	$W_n Tendency_t^{posR} Ratio_t - W_n Tendency_{t-1}^{posR} Ratio_{t-1}$
$W_n Tendency_t^{posR} DiffH_t^m$	$W_n Tendency_t^{posR} - W_m Tendency_{t-n}^{posR}$
$W_n Tendency_t^{posR} RatioDiffH_t^m$	$W_n Tendency_t^{posR} Ratio_t - W_m Tendency_{t-n}^{posR} Ratio_{t-n}$
$W_n Tendency_t^{negR}$	$\sum_{i=0}^N \{1 \text{ if } ReviewPost_i < ReviewNeg_i \text{ else } 0\}$
$W_n Tendency_t^{negR} Diff_t$	$W_n Tendency_t^{negR} - W_n Tendency_{t-1}^{negR}$
$W_n Tendency_t^{negR} DiffRatio_t$	$\frac{W_n Tendency_t^{negR} - W_n Tendency_{t-1}^{negR}}{W_n Tendency_{t-1}^{negR}}$
$W_n Tendency_t^{negR} Ratio_t$	$\frac{W_n Tendency_t^{negR}}{W_n Review_t}$
$W_n Tendency_t^{negR} RatioDiff_t$	$W_n Tendency_t^{negR} Ratio_t - W_n Tendency_{t-1}^{negR} Ratio_{t-1}$
$W_n Tendency_t^{negR} DiffH_t^m$	$W_n Tendency_t^{negR} - W_m Tendency_{t-n}^{negR}$
$W_n Tendency_t^{negR} RatioDiffH_t^m$	$W_n Tendency_t^{negR} Ratio_t - W_m Tendency_{t-n}^{negR} Ratio_{t-n}$
$W_n Tendency_t^{pos}$	$\sum_{i=0}^N \{ReviewTen_i \text{ if } ReviewPost_i > ReviewNeg_i \text{ else } 0\}$
$W_n Tendency_t^{pos} Diff_t$	$W_n Tendency_t^{pos} - W_n Tendency_{t-1}^{pos}$
$W_n Tendency_t^{pos} DiffRatio_t$	$\frac{W_n Tendency_t^{pos} - W_n Tendency_{t-1}^{pos}}{W_n Tendency_{t-1}^{pos}}$
$W_n Tendency_t^{pos} Average_t$	$\frac{W_n Tendency_t^{pos}}{W_n Tendency_t^{posR}}$
$W_n Tendency_t^{pos} AverageDiff_t$	$W_n Tendency_t^{pos} Average_t - W_n Tendency_{t-1}^{pos} Average_{t-1}$
$W_n Tendency_t^{pos} DiffH_t^m$	$W_n Tendency_t^{pos} - W_m Tendency_{t-n}^{pos}$
$W_n Tendency_t^{pos} AverageDiffH_t^m$	$W_n Tendency_t^{pos} Average_t - W_m Tendency_{t-n}^{pos} Average_{t-n}$
$W_n Tendency_t^{neg}$	$\sum_{i=0}^N \{ReviewTen_i \text{ if } ReviewPost_i < ReviewNeg_i \text{ else } 0\}$
$W_n Tendency_t^{neg} Diff_t$	$W_n Tendency_t^{neg} - W_n Tendency_{t-1}^{neg}$
$W_n Tendency_t^{neg} DiffRatio_t$	$\frac{W_n Tendency_t^{neg} - W_n Tendency_{t-1}^{neg}}{W_n Tendency_{t-1}^{neg}}$
$W_n Tendency_t^{neg} Average_t$	$\frac{W_n Tendency_t^{neg}}{W_n Tendency_t^{negR}}$
$W_n Tendency_t^{neg} AverageDiff_t$	$W_n Tendency_t^{neg} Average_t - W_n Tendency_{t-1}^{neg} Average_{t-1}$
$W_n Tendency_t^{neg} DiffH_t^m$	$W_n Tendency_t^{neg} - W_m Tendency_{t-n}^{neg}$

Table A.9 continued from previous page

Feature	Definition
$W_n Tendency^{neg} AverageDiffH_t^m$	$W_n Tendency^{neg} Average_t - W_m Tendency^{neg} Average_{t-n}$
$W_n Tendency_t$	$W_n Tendency_t^{pos} + W_n Tendency_t^{neg}$
$W_n TendencyDiff_t$	$W_n Tendency_t - W_n Tendency_{t-1}$
$W_n TendencyDiffRatio_t$	$\frac{W_n Tendency_t - W_n Tendency_{t-1}}{W_n Tendency_{t-1}}$
$W_n TendencyDiffH_t^m$	$W_n Tendency_t - W_m Tendency_{t-n}$

Table A.2: Variant Features

Feature	Definition
$W_nBasicFeatureDiff_t$	$W_nBasicFeature_t - W_nBasicFeature_{t-1}$
$W_nBasicFeatureDiffRatio_t$	$\frac{W_nBasicFeature_t - W_nBasicFeature_{t-1}}{W_nBasicFeature_{t-1}}$
$W_nBasicFeatureRatio_t$	$\frac{W_nBasicFeature_t}{W_nReview_t}$
$W_nBasicFeatureRatioDiff_t$	$W_nBasicFeatureRatio_t - W_nBasicFeatureRatio_{t-1}$
$W_nBasicFeatureDiffH_t^m$	$W_nBasicFeature_t - W_mBasicFeature_{t-n}$
$W_nBasicFeatureRatioDiffH_t^m$	$W_nBasicFeatureRatio_t - W_mBasicFeatureRatio_{t-n}$
$W_nBasicFeatureAverage_t$	$\frac{W_nBasicFeature_t}{W_nReview_t}$
$W_nBasicFeatureAverageDiff_t$	$W_nBasicFeatureAverage_t - W_nBasicFeatureAverage_{t-1}$
$W_nBasicFeatureAverageDiffH_t^m$	$W_nBasicFeatureAverage_t - W_mBasicFeatureAverage_{t-n}$

The *Diff* variant is the difference sequence of each feature series. The *DiffRatio* variant represents the rate of change between two adjacent time points. The *Ratio* variant is the proportion of reviews with certain features in n weeks. The *RatioDiff* variant represents the proportion change between two adjacent time points. The *DiffH^m* variant is the difference between values of certain features in n weeks and that in the prior m weeks, in which M is the number of reviews in $[t - n - m, t - n]$ weeks, where m is the length of relative history data window and $m \in \{4, 6, 8, 10, 12, 16, 20, 24 | m > n\}$. The *RatioDiffH^m* variant represents the difference between the proportion of reviews with some feature in n weeks and that in the prior m weeks. The formula of *Average* variant is the same with *Ratio* variant, but indicates the average value of reviews in some feature within n weeks, such as $W_nUsefulAverage$ means the average number of useful vote for every review in n weeks. The *AverageDiff* variant shows the average value change between two adjacent time points. The *AverageDiffH^m* variant expresses the difference between the average value of reviews with some feature in n weeks and that in the prior m weeks.

Table A.3: Review Features

Feature	Definition
$W_nReview_t$	N
$W_nReviewDiff_t$	$W_nReview_t - W_nReview_{t-1}$
$W_nReviewDiffRatio_t$	$\frac{W_nReview_t - W_nReview_{t-1}}{W_nReview_{t-1}}$
$W_nReviewDiffH_t^m$	$W_nReview_t - W_mReview_{t-n}$

Table A.4: Star Features

Feature	Definition
$W_nStar_t^s$	$\sum_{i=1}^N \{1 \text{ if } score_i == s \text{ else } 0; s \in [1, 5]\}$
$W_nStar^sDiff_t$	$W_nStar_t^s - W_nStar_{t-1}^s$
$W_nStar^sDiffRatio_t$	$\frac{W_nStar_t^s - W_nStar_{t-1}^s}{W_nStar_{t-1}^s}$
$W_nStar^sRatio_t$	$\frac{W_nStar_t^s}{W_nReview_t}$
$W_nStar^sRatioDiff_t$	$W_nStar^sRatio_t - W_nStar^sRatio_{t-1}$
$W_nStar^sDiffH_t^m$	$W_nStar_t^s - W_mStar_{t-n}^s$
$W_nStar^sRatioDiffH_t^m$	$W_nStar^sRatio_t - W_mStar^sRatio_{t-n}$
$W_nStar^{15}Diff_t$	$W_nStar_t^5 - W_nStar_t^1$
$W_nStar^{15}DiffRatio_t$	$\frac{W_nStar^{15}Diff_t - W_nStar^{15}Diff_{t-1}}{W_nStar^{15}Diff_{t-1}}$
$W_nStar^{15}Ratio_t$	$W_nStar^5Ratio_t - W_nStar^1Ratio_t$
$W_nStar^{15}RatioDiff_t$	$W_nStar^{15}Ratio_t - W_nStar^{15}Ratio_{t-1}$

Table A.5: Default Features

Feature	Definition
$W_nDefault_t$	$\sum_{i=1}^N \{1 \text{ if } isDefault_{ti} == True \text{ else } 0\}$
$W_nDefaultDiff_t$	$W_nDefault_t - W_nDefault_{t-1}$
$W_nDefaultDiffRatio_t$	$\frac{W_nDefault_t - W_nDefault_{t-1}}{W_nDefault_{t-1}}$
$W_nDefaultRatio_t$	$\frac{W_nDefault_t}{W_nReview_t}$
$W_nDefaultRatioDiff_t$	$W_nDefaultRatio_t - W_nDefaultRatio_{t-1}$
$W_nDefaultDiffH_t^m$	$W_nDefault_t - W_mDefault_{t-n}$
$W_nDefaultRatioDiffH_t^m$	$W_nDefaultRatio_t - W_mDefaultRatio_{t-n}$

Table A.6: Score Features

Feature	Definition
W_nScore_t	$\frac{\sum_{s=1}^5 (s \times W_nStar_t^s)}{W_nReview_t}$
$W_nScoreDiff_t$	$W_nScore_t - W_nScore_{t-1}$
$W_nScoreDiffRatio_t$	$\frac{W_nScore_t - W_nScore_{t-1}}{W_nScore_{t-1}}$
$W_nScoreDiffH_t^m$	$W_nScore_t - W_mScore_{t-n}$

Table A.7: Emotion Variant Features

Feature	Definition
$W_nEmotionFeatureRatioE_t$	$\frac{W_nEmotionFeature_t}{W_nEmotion_t}$
$W_nEmotionFeatureRatioEDiff_t$	$W_nEmotionFeatureRatioE_t - W_nEmotionFeatureRatioE_{t-1}$
$W_nEmotionFeatureRatioEDiffH_t^n$	$W_nEmotionFeatureRatioE_t - W_mEmotionFeatureRatioE_{t-n}$

The *RatioE* variant represents the proportion of reviews with certain emotions within emotional reviews in n weeks and the other two have the same meaning with *RatioDiff* and *RatioDiffH^m* however the number of emotional reviews $W_nEmotion$ as denominator. These three special variants are only applicable to $W_nEmotion^e$ and $W_nEmotion^{negative}$.

Table A.8: Emotion Features

Feature	Definition
$W_n Emotion_t^e$	$\sum_{i=1}^N \{1 \text{ if } ReviewEmotion_{ti} == e \text{ else } 0; e \in [0, 4]\}$
$W_n Emotion^e Diff_t$	$W_n Emotion_t^e - W_n Emotion_{t-1}^e$
$W_n Emotion^e DiffRatio_t$	$\frac{W_n Emotion_t^e - W_n Emotion_{t-1}^e}{W_n Emotion_{t-1}^e}$
$W_n Emotion^e Ratio_t$	$\frac{W_n Emotion_t^e}{W_n Review_t}$
$W_n Emotion^e RatioDiff_t$	$W_n Emotion^e Ratio_t - W_n Emotion^e Ratio_{t-1}$
$W_n Emotion^e DiffH_t^m$	$W_n Emotion_t^e - W_m Emotion_{t-n}^e$
$W_n Emotion^e RatioDiffH_t^m$	$W_n Emotion^e Ratio_t - W_m Emotion^e Ratio_{t-n}$
$W_n Emotion_t$	$\sum_{e=0}^4 W_n Emotion_t^e$
$W_n Emotion^e RatioE_t$	$\frac{W_n Emotion_t^e}{W_n Emotion_t}$
$W_n Emotion^e RatioEDiff_t$	$W_n Emotion^e RatioE_t - W_n Emotion^e RatioE_{t-1}$
$W_n Emotion^e RatioEDiffH_t^m$	$W_n Emotion^e RatioE_t - W_m Emotion^e RatioE_{t-n}$
$W_n EmotionDiff_t$	$W_n Emotion_t - W_n Emotion_{t-1}$
$W_n EmotionDiffRatio_t$	$\frac{W_n Emotion_t - W_n Emotion_{t-1}}{W_n Emotion_{t-1}}$
$W_n EmotionRatio_t$	$\frac{W_n Emotion_t}{W_n Review_t}$
$W_n EmotionRatioDiff_t$	$W_n EmotionRatio_t - W_n EmotionRatio_{t-1}$
$W_n EmotionDiffH_t^m$	$W_n Emotion_t - W_m Emotion_{t-n}$
$W_n EmotionRatioDiffH_t^m$	$W_n EmotionRatio_t - W_m EmotionRatio_{t-n}$
$W_n Emotion_t^{negative}$	$\sum_{i=1}^N \{1 \text{ if } ReviewEmotion_{ti} \in \{0, 1, 3, 4\} \text{ else } 0\}$
$W_n Emotion^{negative} Diff_t$	$W_n Emotion_t^{negative} - W_n Emotion_{t-1}^{negative}$
$W_n Emotion^{negative} DiffRatio_t$	$\frac{W_n Emotion_t^{negative} - W_n Emotion_{t-1}^{negative}}{W_n Emotion_{t-1}^{negative}}$
$W_n Emotion^{negative} Ratio_t$	$\frac{W_n Emotion_t^{negative}}{W_n Review_t}$
$W_n Emotion^{negative} RatioDiff_t$	$W_n Emotion^{negative} Ratio_t - W_n Emotion^{negative} Ratio_{t-1}$
$W_n Emotion^{negative} DiffH_t^m$	$W_n Emotion_t^{negative} - W_m Emotion_{t-n}^{negative}$
$W_n Emotion^{negative} RatioDiffH_t^m$	$W_n Emotion^{negative} Ratio_t - W_m Emotion^{negative} Ratio_{t-n}$
$W_n Emotion^{negative} RatioE_t$	$\frac{W_n Emotion_t^{negative}}{W_n Emotion_t}$
$W_n Emotion^{negative} RatioEDiff_t$	$W_n Emotion^{negative} RatioE_t - W_n Emotion^{negative} RatioE_{t-1}$
$W_n Emotion^{negative} RatioEDiffH_t^m$	$W_n Emotion^{negative} RatioE_t - W_m Emotion^{negative} RatioE_{t-n}$

Table A.10: Days Features

Feature	Definition
$W_n Days_t$	$\frac{\sum_{i=0}^N days_{ti}}{W_n Review_t}$
$W_n DaysDiff_t$	$W_n Days_t - W_n Days_{t-1}$
$W_n DaysDiffRatio_t$	$\frac{W_n Days_t - W_n Days_{t-1}}{W_n Days_{t-1}}$
$W_n DaysH_t^m$	$W_n Days_t \times W_n Review_t - W_m Days_{t-n} \times W_m Review_{t-n}$
$W_n DaysDiffH_t^m$	$W_n Days_t - W_m Days_{t-n}$

Table A.11: Useful Features

Feature	Definition
$W_n Useful_t$	$\sum_{i=1}^N usefulVoteCount_{ti}$
$W_n UsefulDiff_t$	$W_n Useful_t - W_n Useful_{t-1}$
$W_n UsefulDiffRatio_t$	$\frac{W_n Useful_t - W_n Useful_{t-1}}{W_n Useful_{t-1}}$
$W_n UsefulAverage_t$	$\frac{W_n Useful_t}{W_n Review_t}$
$W_n UsefulAverageDiff_t$	$W_n UsefulAverage_t - W_n UsefulAverage_{t-1}$
$W_n UsefulDiffH_t^m$	$W_n Useful_t - W_m Useful_{t-n}$
$W_n UsefulAverageDiffH_t^m$	$W_n UsefulAverage_t - W_m UsefulAverage_{t-n}$
$W_n UsefulR_t$	$\sum_{i=1}^N \{1 \text{ if } usefulVoteCount_{ti} > 0 \text{ else } 0\}$
$W_n UsefulRDiff_t$	$W_n UsefulR_t - W_n UsefulR_{t-1}$
$W_n UsefulRDiffRatio_t$	$\frac{W_n UsefulR_t - W_n UsefulR_{t-1}}{W_n UsefulR_{t-1}}$
$W_n UsefulRRatio_t$	$\frac{W_n UsefulR_t}{W_n Review_t}$
$W_n UsefulRRatioDiff_t$	$W_n UsefulRRatio_t - W_n UsefulRRatio_{t-1}$
$W_n UsefulRDiffH_t^m$	$W_n UsefulR_t - W_m UsefulR_{t-n}$
$W_n UsefulRRatioDiffH_t^m$	$W_n UsefulRRatio_t - W_m UsefulRRatio_{t-n}$

Table A.12: Useless Features

Feature	Definition
$W_n Useless_t$	$\sum_{i=1}^N uselessVoteCount_{ti}$
$W_n UselessDiff_t$	$W_n Useless_t - W_n Useless_{t-1}$
$W_n UselessDiffRatio_t$	$\frac{W_n Useless_t - W_n Useless_{t-1}}{W_n Useless_{t-1}}$
$W_n UselessAverage_t$	$\frac{W_n Useless_t}{W_n Review_t}$
$W_n UselessAverageDiff_t$	$W_n UselessAverage_t - W_n UselessAverage_{t-1}$
$W_n UselessDiffH_t^m$	$W_n Useless_t - W_m Useless_{t-n}$
$W_n UselessAverageDiffH_t^m$	$W_n UselessAverage_t - W_m UselessAverage_{t-n}$
$W_n UselessR_t$	$\sum_{i=1}^N \{1 \text{ if } uselessVoteCount_{ti} > 0 \text{ else } 0\}$
$W_n UselessRDiff_t$	$W_n UselessR_t - W_n UselessR_{t-1}$
$W_n UselessRDiffRatio_t$	$\frac{W_n UselessR_t - W_n UselessR_{t-1}}{W_n UselessR_{t-1}}$
$W_n UselessRRatio_t$	$\frac{W_n UselessR_t}{W_n Review_t}$
$W_n UselessRRatioDiff_t$	$W_n UselessRRatio_t - W_n UselessRRatio_{t-1}$
$W_n UselessRDiffH_t^m$	$W_n UselessR_t - W_m UselessR_{t-n}$
$W_n UselessRRatioDiffH_t^m$	$W_n UselessRRatio_t - W_m UselessRRatio_{t-n}$

Table A.13: Image Features

Feature	Definition
$W_n Image_t$	$\sum_{i=1}^N img_{ti}$
$W_n ImageDiff_t$	$W_n Image_t - W_n Image_{t-1}$
$W_n ImageDiffRatio_t$	$\frac{W_n Image_t - W_n Image_{t-1}}{W_n Image_{t-1}}$
$W_n ImageAverage_t$	$\frac{W_n Image_t}{W_n Review_t}$
$W_n ImageAverageDiff_t$	$W_n ImageAverage_t - W_n ImageAverage_{t-1}$
$W_n ImageDiffH_t^m$	$W_n Image_t - W_m Image_{t-n}$
$W_n ImageAverageDiffH_t^m$	$W_n ImageAverage_t - W_m ImageAverage_{t-n}$
$W_n ImageR_t$	$\sum_{i=1}^N \{1 \text{ if } img_{ti} > 0 \text{ else } 0\}$
$W_n ImageRDiff_t$	$W_n ImageR_t - W_n ImageR_{t-1}$
$W_n ImageRDiffRatio_t$	$\frac{W_n ImageR_t - W_n ImageR_{t-1}}{W_n ImageR_{t-1}}$
$W_n ImageRRatio_t$	$\frac{W_n ImageR_t}{W_n Review_t}$
$W_n ImageRRatioDiff_t$	$W_n ImageRRatio_t - W_n ImageRRatio_{t-1}$
$W_n ImageRDiffH_t^m$	$W_n ImageR_t - W_m ImageR_{t-n}$
$W_n ImageRRatioDiffH_t^m$	$W_n ImageRRatio_t - W_m ImageRRatio_{t-n}$

Table A.14: Reply Features

Feature	Definition
$W_n Reply_t$	$\sum_{i=1}^N replyCount_{ti}$
$W_n ReplyDiff_t$	$W_n Reply_t - W_n Reply_{t-1}$
$W_n ReplyDiffRatio_t$	$\frac{W_n Reply_t - W_n Reply_{t-1}}{W_n Reply_{t-1}}$
$W_n ReplyAverage_t$	$\frac{W_n Reply_t}{W_n Review_t}$
$W_n ReplyAverageDiff_t$	$W_n ReplyAverage_t - W_n ReplyAverage_{t-1}$
$W_n ReplyDiffH_t^m$	$W_n Reply_t - W_m Reply_{t-n}$
$W_n ReplyAverageDiffH_t^m$	$W_n ReplyAverage_t - W_m ReplyAverage_{t-n}$
$W_n ReplyR_t$	$\sum_{i=1}^N \{1 \text{ if } replyCount_{ti} > 0 \text{ else } 0\}$
$W_n ReplyRDiff_t$	$W_n ReplyR_t - W_n ReplyR_{t-1}$
$W_n ReplyRDiffRatio_t$	$\frac{W_n ReplyR_t - W_n ReplyR_{t-1}}{W_n ReplyR_{t-1}}$
$W_n ReplyRRatio_t$	$\frac{W_n ReplyR_t}{W_n Review_t}$
$W_n ReplyRRatioDiff_t$	$W_n ReplyRRatio_t - W_n ReplyRRatio_{t-1}$
$W_n ReplyRDiffH_t^m$	$W_n ReplyR_t - W_m ReplyR_{t-n}$
$W_n ReplyRRatioDiffH_t^m$	$W_n ReplyRRatio_t - W_m ImageRRatio_{t-n}$

Table A.15: Client Features

Feature	Definition
$W_n Client_t^c$	$\sum_{i=1}^N \{1 \text{ if } userClient_{ti} == c \text{ else } 0\}$
$W_n Client^c Diff_t$	$W_n Client_t^c - W_n Client_{t-1}^c$
$W_n Client^c DiffRatio_t$	$\frac{W_n Client_t^c - W_n Client_{t-1}^c}{W_n Client_{t-1}^c}$
$W_n Client^c Ratio_t$	$\frac{W_n Client_t^c}{W_n Review_t}$
$W_n Client^c RatioDiff_t$	$W_n Client^c Ratio_t - W_n Client^c Ratio_{t-1}$
$W_n Client^c DiffH_t^m$	$W_n Client_t^c - W_m Client_{t-n}^c$
$W_n Client^c RatioDiffH_t^m$	$W_n Client^c Ratio_t - W_m Client^c Ratio_{t-n}$

Table A.16: Mobile Features

Feature	Definition
$W_n Mobile_t$	$\sum_{i=1}^N \{1 \text{ if } isMobile_{ti} == True \text{ else } 0\}$
$W_n MobileDiff_t$	$W_n Mobile_t - W_n Mobile_{t-1}$
$W_n MobileDiffRatio_t$	$\frac{W_n Mobile_t - W_n Mobile_{t-1}}{W_n Mobile_{t-1}}$
$W_n MobileRatio_t$	$\frac{W_n Mobile_t}{W_n Review_t}$
$W_n MobileRatioDiff_t$	$W_n MobileRatio_t - W_n MobileRatio_{t-1}$
$W_n MobileDiffH_t^m$	$W_n Mobile_t - W_m Mobile_{t-n}$
$W_n MobileRatioDiffH_t^m$	$W_n MobileRatio_t - W_m MobileRatio_{t-n}$

Table A.17: Average accuracies of classifiers with different feature categories by 5-fold cross-validations on the training set.

	XGB	GradientBoosting	AdaBoost	LSTM	LogisticRegression	Bagging	RandomForest	GaussianNB
<i>Known</i>	56.75%	56.41%	56.48%	55.67%	56.06%	53.08%	51.53%	50.31%
Review	54.44%	54.35%	54.28%	54.45%	52.02%	51.39%	51.00%	55.30%
Star	53.83%	53.99%	52.92%	55.49%	52.84%	48.92%	49.02%	52.61%
Default	55.33%	55.17%	55.34%	55.74%	56.43%	55.36%	54.44%	43.14%
Score	54.85%	54.92%	55.30%	54.41%	56.47%	50.72%	50.84%	51.96%
Emotion	57.36%	57.16%	57.39%	55.70%	54.73%	53.42%	52.51%	55.39%
Tendency	53.50%	53.74%	53.55%	54.48%	55.52%	49.20%	47.86%	44.57%
<i>Consumer Experience</i>	54.83%	54.79%	54.02%	57.07%	55.73%	51.47%	51.96%	47.00%
Days	54.17%	54.27%	53.79%	54.47%	55.72%	51.82%	50.82%	46.04%
Useful	54.20%	54.70%	54.27%	57.71%	52.23%	50.93%	51.41%	46.62%
Useless	56.50%	56.22%	55.88%	56.87%	56.69%	54.89%	54.87%	46.48%
Image	55.61%	55.07%	54.75%	56.86%	55.80%	51.13%	51.49%	46.01%
Reply	51.98%	51.93%	50.80%	56.76%	52.74%	48.86%	48.32%	46.53%
<i>Consumer</i>	58.74%	58.76%	57.07%	59.33%	55.15%	54.96%	54.75%	55.21%
Client	58.39%	58.25%	57.69%	57.38%	54.70%	54.99%	54.04%	55.06%
Mobile	56.99%	56.97%	55.30%	55.41%	52.62%	54.64%	53.90%	52.31%

Table A.18: Average accuracies of classifiers with 1 to 12 weeks statistical granularity by 5-fold cross-validations on the training set.

	XGB	GradientBoosting	AdaBoost	LSTM	LogisticRegression	Bagging	RandomForest	GaussianNB
1	54.64%	54.56%	54.58%	55.05%	54.79%	51.65%	50.40%	48.84%
2	55.91%	55.39%	54.16%	55.54%	54.96%	52.68%	51.21%	48.91%
3	55.81%	55.99%	55.46%	56.17%	55.06%	52.81%	51.48%	49.10%
4	56.73%	57.07%	55.60%	56.15%	55.32%	52.68%	52.57%	49.07%
5	57.05%	56.79%	54.94%	56.48%	54.96%	53.32%	51.74%	49.53%
6	57.13%	56.77%	55.43%	57.03%	55.00%	54.27%	50.98%	49.26%
7	58.80%	58.51%	56.85%	59.41%	55.20%	53.62%	53.04%	49.26%
8	59.36%	58.74%	57.44%	59.01%	55.05%	53.74%	53.19%	49.55%
9	59.27%	59.13%	57.40%	59.41%	54.65%	53.14%	53.51%	52.72%
10	59.24%	59.63%	57.58%	59.34%	54.81%	54.32%	52.32%	52.68%
11	59.85%	59.53%	58.68%	59.80%	55.05%	54.25%	53.08%	52.58%
12	57.77%	58.03%	58.10%	59.44%	55.24%	54.33%	52.72%	52.36%

Table A.19: Ten technical indicators and their definitions

Name of Indicators	Definition
Simple n(10 here)-week Moving Average	$\frac{C_t + C_{t-1} + \dots + C_{t-9}}{n}$
Weighted n(10 here)-week Moving Average	$\frac{10 \times C_t + 9 \times C_{t-1} + \dots + C_{t-9}}{n + (n-1) + \dots + 1}$
Momentum	$C_t - C_{t-9}$
Stochastic K%	$\frac{C_t - LL_{t-(n-1)}}{HH_{t-(n-1)} - LL_{t-(n-1)}} \times 100$
Stochastic D%	$\frac{\sum_{i=0}^{n-1} K_{t-1} \%}{10}$
Relative Strength Index(RSI)	$100 - \frac{100}{1 + (\sum_{i=0}^{n-1} UP_{t-i}/n) / (\sum_{i=0}^{n-1} DW_{t-i}/n)}$
Moving Average Convergence Divergence(MACD)	$MACD(n)_{t-1} + \frac{2}{n+1} \times (DIFF_t - MACD(n)_{t-1})$
Larry William's R%	$\frac{H_n - C_t}{H_n - L_n} \times 100$
A/D (Accumulation/Distribution) Oscillator	$\frac{H_t - C_{t-1}}{H_t - L_t}$
CCI (Commodity Channel Index)	$\frac{M_t - SM_t}{0.015 D_t}$

C_t is the weekly close price i.e. $CLOSE_{W_t}$, L_t is the weekly low price, i.e., LOW_{W_t} and H_t is the weekly high price, i.e., $HIGH_{W_t}$ at week t . $DIFF_t = EMA(12)_t - EMA(26)_t$, EMA is the exponential moving average, $EMA(k)_t = EMA(k)_{t-1} + \alpha \times (C_t - EMA(k)_{t-1})$, α is a smoothing factor which is equal to $\frac{2}{k+1}$, k is the time period of k -week exponential moving average. LL_t and HH_t implies lowest low and highest high in last t weeks, respectively. $M_t = \frac{H_t + L_t + C_t}{3}$, $SM_t = \frac{\sum_{i=1}^n M_{t-i+1}}{n}$, $D_t = \frac{\sum_{i=1}^n |M_{t-i+1} - SM_t|}{n}$. UP_t means upward price change while DW_t is the downward price change at week t .