**A Corpus of Truthful and Deceptive Hotel Reviews**

**By: Jennifer Adrianne I. Caballes and Christian Ted Ochoa**

**Dataset Description**

The dataset is a collection of truthful and deceptive reviews for 20 hotels in Chicago. It is taken from Kaggle, posted by Rachel Tatman [1]. The dataset contains five (5) different columns namely;

1. *Deceptive*: Classifies whether the review is truthful or deceptive.
2. *Hotel:* The column that contains the name of the hotel given the reviews.
3. *Polarity:* Classifies whether the sentiments are positive or negative.
4. *Source:* Contains the source of information.
5. *Text Review:* The actual review posted for the particular hotel.

Hence, A total of 1600 data is present and 80% will be used for training while the other 20% will later be used for testing. Thus, for this project the proponents uses the ‘Text Review’ as the only feature to classify whether it is truthful and deceptive.

**Frequency Matrix**

|  |  |  |
| --- | --- | --- |
|  | **TRUTHFUL** | **DECEPTIVE** |
| **1** | 49 | 15 |
| **$2** | 9 | 0 |
| **4** | 69 | 14 |
| **6** | 30 | 13 |
| **10** | 37 | 15 |
| **12** | 9 | 1 |
| **13** | 6 | 3 |
| **15** | 29 | 15 |
| **20** | 27 | 9 |
| **30** | 27 | 17 |
| **$100** | 9 | 1 |
| **3rd** | 11 | 0 |
| **5th** | 8 | 1 |
| **a** | 613 | 586 |
| **able** | 41 | 49 |
| **absolutely** | 25 | 37 |
| **accommodating** | 7 | 12 |
| **accommodations** | 4 | 40 |
| **across** | 24 | 12 |
| **acted** | 2 | 8 |
| **advance** | 12 | 6 |
| **advantage** | 4 | 13 |
| **affinia** | 14 | 29 |
| **after** | 143 | 148 |
| **again** | 174 | 181 |
| **against** | 7 | 3 |
| **allegro** | 21 | 16 |
| **allow** | 7 | 11 |
| **almost** | 36 | 27 |
| **along** | 16 | 19 |
| **also** | 140 | 119 |
| **am** | 72 | 48 |
| **amazed** | 1 | 8 |
| **amenities** | 32 | 59 |
| **amenity** | 2 | 7 |
| **amount** | 6 | 12 |
| **anniversary** | 7 | 27 |
| **another** | 71 | 59 |
| **answer** | 8 | 3 |
| **appetizers** | 6 | 3 |
| **appointed** | 20 | 7 |
| **aquarium** | 3 | 7 |
| **areas** | 17 | 12 |
| **arrived** | 73 | 80 |
| **as** | 230 | 242 |
| **ask** | 20 | 25 |
| **asked** | 74 | 55 |
| **atmosphere** | 10 | 36 |
| **attend** | 3 | 7 |
| **attended** | 9 | 2 |
| **attractions** | 9 | 26 |
| **available** | 39 | 51 |
| **aveda** | 10 | 2 |
| **average** | 19 | 15 |
| **bad** | 55 | 43 |
| **bag** | 6 | 3 |
| **bags** | 13 | 23 |
| **bargain** | 7 | 2 |
| **based** | 15 | 5 |
| **basically** | 11 | 6 |
| **bath** | 18 | 11 |
| **bathrooms** | 27 | 10 |
| **be** | 266 | 278 |
| **beat** | 13 | 6 |
| **beautiful** | 54 | 87 |
| **because** | 104 | 93 |
| **before** | 76 | 66 |
| **bellhop** | 3 | 6 |
| **below** | 18 | 2 |
| **better** | 94 | 77 |
| **between** | 19 | 9 |
| **big** | 44 | 37 |
| **bill** | 21 | 8 |
| **bit** | 50 | 32 |
| **blanket** | 7 | 2 |
| **block** | 25 | 2 |
| **blocks** | 31 | 4 |
| **bother** | 10 | 7 |
| **bottles** | 11 | 1 |
| **bottom** | 9 | 3 |
| **boutique** | 15 | 4 |
| **bowl** | 8 | 3 |
| **breakfast** | 93 | 57 |
| **breath** | 1 | 11 |
| **brick** | 8 | 2 |
| **bright** | 7 | 2 |
| **broken** | 25 | 9 |
| **brought** | 20 | 23 |
| **brown** | 6 | 3 |
| **bug** | 6 | 3 |
| **built** | 6 | 3 |
| **but** | 371 | 254 |
| **c** | 10 | 7 |
| **call** | 65 | 27 |
| **called** | 97 | 52 |
| **calling** | 13 | 4 |
| **calls** | 17 | 9 |
| **ceiling** | 7 | 2 |
| **celebrate** | 1 | 8 |
| **centre** | 8 | 1 |
| **chair** | 17 | 3 |
| **charge** | 47 | 14 |
| **charges** | 15 | 3 |
| **charging** | 10 | 0 |
| **charm** | 12 | 5 |
| **cheaper** | 7 | 2 |
| **checkout** | 14 | 5 |
| **children** | 6 | 13 |
| **choose** | 21 | 13 |
| **chosen** | 7 | 3 |
| **cigarette** | 1 | 11 |
| **city** | 62 | 110 |
| **classy** | 3 | 17 |
| **cleaned** | 14 | 32 |
| **cleaning** | 10 | 24 |
| **cleanliness** | 3 | 15 |
| **close** | 65 | 53 |
| **closed** | 11 | 6 |
| **closer** | 8 | 1 |
| **closet** | 25 | 4 |
| **clothes** | 7 | 2 |
| **cocktails** | 6 | 3 |
| **com** | 8 | 1 |
| **complaint** | 15 | 4 |
| **complaints** | 10 | 7 |
| **concern** | 6 | 3 |
| **concierge** | 59 | 33 |
| **confirm** | 7 | 2 |
| **confirmed** | 9 | 2 |
| **connection** | 3 | 7 |
| **conrad** | 13 | 31 |
| **consider** | 8 | 2 |
| **continue** | 7 | 2 |
| **conveniently** | 6 | 14 |
| **correct** | 12 | 5 |
| **costs** | 9 | 3 |
| **couch** | 8 | 2 |
| **could** | 140 | 135 |
| **couldn** | 38 | 51 |
| **course** | 20 | 9 |
| **courteous** | 10 | 16 |
| **cover** | 8 | 2 |
| **covered** | 17 | 2 |
| **cramped** | 9 | 1 |
| **credit** | 19 | 9 |
| **cup** | 10 | 1 |
| **currently** | 8 | 1 |
| **cut** | 8 | 1 |
| **daily** | 8 | 3 |
| **dark** | 22 | 7 |
| **dated** | 15 | 2 |
| **daughter** | 13 | 6 |
| **deal** | 40 | 22 |
| **decent** | 22 | 14 |
| **decided** | 27 | 62 |
| **delicious** | 5 | 32 |
| **delight** | 3 | 7 |
| **delivered** | 18 | 10 |
| **deluxe** | 5 | 12 |
| **desk** | 157 | 133 |
| **despite** | 21 | 15 |
| **didn** | 94 | 88 |
| **didnt** | 11 | 6 |
| **die** | 1 | 11 |
| **difficult** | 18 | 9 |
| **dining** | 11 | 34 |
| **dinner** | 30 | 43 |
| **discover** | 7 | 2 |
| **distance** | 37 | 25 |
| **doesn** | 14 | 6 |
| **dog** | 9 | 18 |
| **dollars** | 13 | 5 |
| **don** | 69 | 39 |
| **door** | 89 | 50 |
| **doorman** | 15 | 2 |
| **doormen** | 18 | 2 |
| **doors** | 22 | 12 |
| **down** | 99 | 73 |
| **drive** | 12 | 6 |
| **e** | 11 | 1 |
| **each** | 34 | 28 |
| **earlier** | 10 | 7 |
| **ease** | 2 | 9 |
| **east** | 21 | 31 |
| **efficient** | 7 | 3 |
| **el** | 10 | 0 |
| **elegant** | 16 | 37 |
| **elevator** | 46 | 8 |
| **elevators** | 40 | 4 |
| **else** | 27 | 34 |
| **ended** | 19 | 16 |
| **enjoy** | 15 | 30 |
| **enjoyable** | 3 | 14 |
| **enjoyed** | 31 | 67 |
| **enough** | 45 | 38 |
| **entered** | 6 | 14 |
| **entire** | 24 | 20 |
| **entrance** | 11 | 6 |
| **even** | 162 | 142 |
| **evening** | 31 | 14 |
| **everyone** | 24 | 31 |
| **everything** | 64 | 85 |
| **exactly** | 7 | 10 |
| **excellent** | 60 | 50 |
| **exercise** | 3 | 6 |
| **expecting** | 11 | 16 |
| **expensive** | 34 | 18 |
| **extremely** | 42 | 48 |
| **eye** | 3 | 6 |
| **facing** | 9 | 0 |
| **fair** | 8 | 3 |
| **famous** | 3 | 7 |
| **feeling** | 11 | 15 |
| **fi** | 11 | 17 |
| **finally** | 45 | 75 |
| **flight** | 7 | 13 |
| **floors** | 24 | 2 |
| **flowers** | 8 | 1 |
| **following** | 8 | 1 |
| **food** | 49 | 104 |
| **forget** | 3 | 8 |
| **forward** | 15 | 28 |
| **four** | 31 | 21 |
| **friend** | 11 | 17 |
| **from** | 286 | 254 |
| **front** | 132 | 114 |
| **fruit** | 9 | 2 |
| **full** | 40 | 24 |
| **fun** | 18 | 27 |
| **funky** | 8 | 1 |
| **furnishings** | 7 | 11 |
| **furniture** | 24 | 20 |
| **given** | 40 | 31 |
| **giving** | 8 | 2 |
| **gold** | 13 | 6 |
| **got** | 127 | 96 |
| **grand** | 1 | 11 |
| **great** | 241 | 197 |
| **group** | 12 | 7 |
| **guests** | 52 | 40 |
| **hall** | 14 | 6 |
| **happen** | 7 | 2 |
| **hardly** | 3 | 8 |
| **hassle** | 3 | 6 |
| **hate** | 2 | 7 |
| **he** | 60 | 50 |
| **headed** | 3 | 8 |
| **hear** | 41 | 12 |
| **heart** | 7 | 28 |
| **heat** | 7 | 2 |
| **helped** | 8 | 18 |
| **helping** | 2 | 7 |
| **high** | 48 | 53 |
| **hilton** | 44 | 56 |
| **his** | 27 | 17 |
| **historic** | 13 | 5 |
| **hold** | 14 | 4 |
| **home** | 34 | 65 |
| **homewood** | 5 | 29 |
| **honeymoon** | 3 | 7 |
| **honor** | 11 | 0 |
| **hopes** | 2 | 7 |
| **hotwire** | 10 | 0 |
| **hour** | 50 | 60 |
| **housekeeping** | 37 | 17 |
| **i** | 492 | 581 |
| **ice** | 14 | 3 |
| **impossible** | 9 | 3 |
| **including** | 25 | 11 |
| **incredible** | 7 | 13 |
| **incredibly** | 17 | 9 |
| **indoor** | 3 | 25 |
| **inside** | 14 | 20 |
| **instead** | 28 | 15 |
| **is** | 418 | 359 |
| **issues** | 18 | 8 |
| **it** | 464 | 448 |
| **itself** | 27 | 35 |
| **james** | 25 | 18 |
| **john** | 10 |  |
| **june** | 9 | 1 |
| **kids** | 11 | 24 |
| **king** | 53 | 46 |
| **known** | 2 | 7 |
| **l** | 14 | 3 |
| **laptop** | 3 | 9 |
| **large** | 62 | 28 |
| **lcd** | 3 | 6 |
| **leaving** | 15 | 11 |
| **left** | 54 | 45 |
| **leg** | 3 | 6 |
| **leisure** | 6 | 3 |
| **letter** | 8 | 3 |
| **life** | 8 | 21 |
| **light** | 18 | 9 |
| **lighting** | 7 | 3 |
| **line** | 50 | 12 |
| **little** | 64 | 35 |
| **live** | 17 | 12 |
| **lobby** | 131 | 75 |
| **local** | 4 | 16 |
| **located** | 40 | 58 |
| **location** | 231 | 104 |
| **locations** | 7 | 3 |
| **locked** | 7 | 2 |
| **lockwood** | 3 | 7 |
| **long** | 48 | 63 |
| **looked** | 43 | 48 |
| **looking** | 51 | 87 |
| **lost** | 18 | 10 |
| **lots** | 23 | 11 |
| **loud** | 18 | 16 |
| **loved** | 41 | 47 |
| **low** | 21 | 5 |
| **lower** | 9 | 2 |
| **lunch** | 7 | 3 |
| **luxurious** | 8 | 54 |
| **luxury** | 7 | 66 |
| **maid** | 11 | 17 |
| **main** | 26 | 10 |
| **maintenance** | 12 | 6 |
| **major** | 19 | 8 |
| **maker** | 9 | 2 |
| **marriott** | 9 | 2 |
| **may** | 29 | 15 |
| **me** | 174 | 199 |
| **meeting** | 13 | 24 |
| **men** | 7 | 3 |
| **mention** | 8 | 20 |
| **mentioned** | 13 | 5 |
| **met** | 9 | 20 |
| **michigan** | 74 | 34 |
| **mid** | 9 | 3 |
| **midway** | 8 | 2 |
| **mile** | 46 | 34 |
| **millenium** | 15 | 4 |
| **mini** | 21 | 13 |
| **mins** | 8 | 2 |
| **missed** | 7 | 2 |
| **modern** | 35 | 57 |
| **moment** | 9 | 20 |
| **monaco** | 22 | 15 |
| **more** | 123 | 136 |
| **morning** | 86 | 45 |
| **most** | 60 | 80 |
| **move** | 18 | 11 |
| **movie** | 8 | 3 |
| **music** | 16 | 10 |
| **navy** | 16 | 21 |
| **near** | 30 | 24 |
| **neat** | 3 | 9 |
| **need** | 55 | 45 |
| **needed** | 27 | 35 |
| **needless** | 6 | 12 |
| **negative** | 12 | 7 |
| **neighborhood** | 15 | 5 |
| **neither** | 9 | 3 |
| **nice** | 172 | 133 |
| **nicely** | 10 | 7 |
| **nicest** | 3 | 6 |
| **night** | 197 | 135 |
| **nightmare** | 2 | 7 |
| **nights** | 86 | 52 |
| **no** | 210 | 109 |
| **noisy** | 24 | 3 |
| **north** | 20 | 8 |
| **not** | 368 | 278 |
| **nothing** | 56 | 52 |
| **number** | 15 | 5 |
| **obviously** | 11 | 15 |
| **odor** | 5 | 14 |
| **off** | 87 | 55 |
| **offer** | 28 | 42 |
| **offered** | 28 | 33 |
| **offers** | 7 | 12 |
| **oh** | 21 | 6 |
| **ok** | 29 | 5 |
| **one** | 212 | 203 |
| **online** | 16 | 20 |
| **only** | 150 | 122 |
| **onto** | 6 | 3 |
| **open** | 51 | 11 |
| **opened** | 17 | 10 |
| **opening** | 8 | 1 |
| **ordered** | 20 | 24 |
| **others** | 20 | 9 |
| **our** | 254 | 211 |
| **outdated** | 6 | 3 |
| **outside** | 33 | 29 |
| **over** | 101 | 82 |
| **overall** | 45 | 74 |
| **own** | 15 | 28 |
| **paid** | 54 | 25 |
| **paint** | 9 | 3 |
| **palmer** | 21 | 16 |
| **pampered** | 9 | 0 |
| **paper** | 19 | 10 |
| **particularly** | 6 | 12 |
| **past** | 23 | 13 |
| **pay** | 42 | 38 |
| **peeling** | 11 | 0 |
| **peninsula** | 10 | maids |
| **people** | 69 | 53 |
| **per** | 28 | 15 |
| **perfect** | 37 | 63 |
| **perhaps** | 13 | 4 |
| **person** | 35 | 26 |
| **personal** | 12 | 22 |
| **pet** | 6 | 20 |
| **pier** | 15 | 21 |
| **pillow** | 8 | 18 |
| **pillows** | 30 | 22 |
| **placed** | 11 | 6 |
| **planned** | 7 | 11 |
| **please** | 13 | 7 |
| **pm** | 15 | 3 |
| **polite** | 11 | 18 |
| **poor** | 27 | 17 |
| **possible** | 11 | 7 |
| **pre** | 7 | 2 |
| **previous** | 13 | 5 |
| **previously** | 6 | 3 |
| **price** | 85 | 66 |
| **prices** | 12 | 22 |
| **problem** | 48 | 23 |
| **process** | 7 | 11 |
| **professional** | 8 | 21 |
| **program** | 7 | 2 |
| **public** | 17 | 2 |
| **put** | 44 | 19 |
| **queen** | 11 | 17 |
| **questions** | 3 | 6 |
| **quick** | 12 | 23 |
| **quite** | 49 | 30 |
| **radio** | 8 | 1 |
| **rate** | 62 | 19 |
| **reach** | 6 | 3 |
| **ready** | 16 | 27 |
| **really** | 91 | 107 |
| **reasonable** | 15 | 22 |
| **reasonably** | 6 | 3 |
| **recent** | 5 | 21 |
| **recommended** | 16 | 19 |
| **record** | 7 | 2 |
| **regency** | 6 | 31 |
| **regular** | 11 | 6 |
| **relatively** | 3 | 6 |
| **relax** | 2 | 32 |
| **relaxed** | 3 | 9 |
| **removed** | 10 | 0 |
| **renovated** | 13 | 6 |
| **renovations** | 9 | 3 |
| **replaced** | 7 | 2 |
| **request** | 26 | 10 |
| **requested** | 38 | 23 |
| **reservations** | 17 | 28 |
| **respect** | 3 | 7 |
| **response** | 9 | 1 |
| **restaurant** | 69 | 84 |
| **restaurants** | 55 | 37 |
| **review** | 22 | 7 |
| **ridiculous** | 15 | 11 |
| **river** | 32 | 12 |
| **rock** | 30 | 41 |
| **romantic** | 8 | 18 |
| **room** | 484 | 441 |
| **roomy** | 8 | 1 |
| **royalty** | 1 | 11 |
| **rude** | 40 | 51 |
| **running** | 14 | 5 |
| **rush** | 11 | 6 |
| **safe** | 13 | 7 |
| **saturday** | 17 | 3 |
| **say** | 75 | 89 |
| **saying** | 14 | 4 |
| **seasons** | 12 | 0 |
| **second** | 46 | 16 |
| **security** | 18 | 1 |
| **seems** | 15 | 20 |
| **seen** | 11 | 16 |
| **selection** | 11 | 7 |
| **self** | 8 | 2 |
| **service** | 217 | 254 |
| **set** | 29 | 17 |
| **several** | 42 | 31 |
| **severely** | 3 | 6 |
| **sheraton** | 12 | 33 |
| **shocked** | 6 | 3 |
| **shops** | 14 | 6 |
| **short** | 28 | 33 |
| **show** | 23 | 11 |
| **showed** | 11 | 7 |
| **showers** | 8 | 2 |
| **side** | 27 | 10 |
| **since** | 45 | 37 |
| **sitting** | 13 | 5 |
| **situation** | 16 | 10 |
| **skyline** | 2 | 10 |
| **sleek** | 2 | 8 |
| **slightly** | 9 | 1 |
| **slippers** | 7 | 3 |
| **small** | 118 | 31 |
| **smell** | 8 | 37 |
| **smoke** | 10 | 27 |
| **smoking** | 9 | 25 |
| **so** | 235 | 205 |
| **soda** | 8 | 1 |
| **someone** | 39 | 33 |
| **something** | 43 | 36 |
| **somewhere** | 11 | 16 |
| **soon** | 10 | 27 |
| **sort** | 7 | 12 |
| **special** | 25 | 36 |
| **spectacular** | 3 | 9 |
| **spending** | 6 | 11 |
| **spent** | 29 | 32 |
| **spoke** | 9 | 1 |
| **spotless** | 2 | 7 |
| **spots** | 2 | 8 |
| **st** | 7 | 3 |
| **staff** | 269 | 267 |
| **stairs** | 8 | 2 |
| **stand** | 9 | 1 |
| **starbucks** | 11 | 7 |
| **start** | 21 | 14 |
| **started** | 10 | 18 |
| **stayed** | 224 | 218 |
| **staying** | 72 | 123 |
| **steakhouse** | 3 | 7 |
| **step** | 3 | 8 |
| **still** | 60 | 29 |
| **stood** | 3 | 6 |
| **store** | 9 | 3 |
| **street** | 62 | 11 |
| **stuff** | 11 | 6 |
| **suggest** | 1 | 9 |
| **suite** | 53 | 55 |
| **suites** | 10 | 45 |
| **superb** | 7 | 11 |
| **surely** | 1 | 8 |
| **surprise** | 9 | 3 |
| **switched** | 7 | 3 |
| **t** | 263 | 240 |
| **table** | 15 | 5 |
| **taken** | 16 | 11 |
| **takes** | 3 | 6 |
| **taste** | 5 | 12 |
| **tasted** | 3 | 7 |
| **tell** | 28 | 18 |
| **telling** | 10 | 1 |
| **temperature** | 7 | 3 |
| **terrible** | 14 | 22 |
| **them** | 84 | 78 |
| **then** | 91 | 48 |
| **think** | 62 | 39 |
| **third** | 10 | 2 |
| **this** | 403 | 380 |
| **though** | 62 | 47 |
| **three** | 41 | 22 |
| **throughout** | 23 | 12 |
| **thru** | 10 | 0 |
| **time** | 145 | 191 |
| **tip** | 15 | 5 |
| **today** | 10 | 0 |
| **toilet** | 28 | 16 |
| **toiletries** | 7 | 2 |
| **told** | 93 | 47 |
| **too** | 67 | 61 |
| **took** | 63 | 76 |
| **top** | 52 | 37 |
| **touches** | 9 | 1 |
| **tower** | 40 | 15 |
| **train** | 25 | 2 |
| **trash** | 2 | 7 |
| **travel** | 40 | 31 |
| **traveling** | 12 | 22 |
| **tub** | 30 | 13 |
| **twice** | 15 | 11 |
| **two** | 116 | 77 |
| **type** | 11 | 15 |
| **unable** | 10 | 7 |
| **uncomfortable** | 19 | 8 |
| **up** | 178 | 142 |
| **updated** | 10 | 1 |
| **upgraded** | 23 | 6 |
| **upon** | 37 | 53 |
| **upper** | 6 | 3 |
| **upscale** | 9 | 19 |
| **upset** | 7 | 3 |
| **use** | 60 | 0 |
| **used** | 37 | 27 |
| **valet** | 32 | 20 |
| **value** | 23 | 3 |
| **ve** | 63 | 39 |
| **very** | 341 | 293 |
| **via** | 8 | 3 |
| **view** | 85 | 78 |
| **visited** | 6 | 21 |
| **visiting** | 17 | 38 |
| **wait** | 35 | 57 |
| **walked** | 26 | 36 |
| **wall** | 36 | 10 |
| **walls** | 34 | 18 |
| **want** | 55 | 65 |
| **wanted** | 26 | 54 |
| **warm** | 18 | 28 |
| **was** | 563 | 542 |
| **waste** | 7 | 10 |
| **water** | 83 | 46 |
| **we** | 374 | 267 |
| **wedding** | 11 | 23 |
| **weekend** | 66 | 86 |
| **well** | 129 | 119 |
| **whenever** | 9 |  |
| **whether** | 2 | 10 |
| **while** | 77 | 109 |
| **wi** | 11 | 17 |
| **will** | 157 | 201 |
| **willing** | 3 | 9 |
| **windows** | 31 | 5 |
| **wine** | 37 | 26 |
| **within** | 41 | 21 |
| **without** | 40 | 23 |
| **woke** | 16 | 10 |
| **wood** | 3 | 6 |
| **working** | 25 | 20 |
| **worn** | 13 | 5 |
| **worst** | 24 | 29 |
| **wouldn** | 33 | 13 |
| **wow** | 7 | 10 |
| **year** | 30 | 22 |
| **years** | 33 | 12 |
| **you** | 276 | 230 |
| **your** | 91 | 84 |
|  | 1 | 10 |

**Results**

There are 635 words extracted from the training datasets. Hence, with the use of these data the proponents were able to classify the test review sets as truthful or deceptive with the use of the Naïve Bayes Classifier.

To determine the accuracy of the classifier the confusion matrix was used. The table below shows the results of the testing phase where in the 320 lines of data was processed.

|  |  |  |  |
| --- | --- | --- | --- |
|  | TRUTHFUL | DECEPTIVE | TOTAL |
| TRUTHFUL | 100 | 60 | **160** |
| DECEPTIVE | 126 | 34 | **160** |
| TOTAL | **226** | **94** | **320** |

As a conclusion with the use of the confusion matrix presented above the model used in has gained an **70.63% accuracy level** and a **miscalculation rate of 29.37%.** Additionally, for the truthful data a rate of **62.50% is given for the True Positive** Category, while a **37.50% is given to the false positive**. Thus, for the deceptive data a **78.75% is considered True Positive** while **21.25% is considered false positive**.

**Insights**

For text classification using NLP, This is a simple (naive) classification method based on Bayes rule. It relies on a very simple representation of the document (called the **bag of words** representation). For this project the proponents, wanted to know whether the review was **truthful** or **deceptive**. The process was to count the number of times each of those words appears in the document, in order to classify the document as **truthful** or **deceptive**.

While using the NLP, the chi-square method as a feature selection was also used. The main advantages for using feature selection algorithms are the facts that it reduces the dimension of our data, it makes the training faster and it can improve accuracy by removing noisy features. As a consequence feature selection can help us to avoid overfitting. Lastly to test the accuracy of the performance of the classification model the confusion matrix is used. The confusion matrix is a method where in it is able to represent the performance classifier, and determine the number of times it was confused.

**References**

Tatman, R. (2017, July 18). Deceptive Opinion Spam Corpus. Retrieved March 17, 2018,

from <https://www.kaggle.com/rtatman/deceptive-opinion-spam-corpus/discussion>

Deceptive Opinion Spam Corpus Myle Ott. (n.d.). Retrieved March 17, 2018,

from <http://myleott.com/op-spam.html>

Markham, K. (2016, June 08). Simple guide to confusion matrix terminology. Retrieved March

21, 2018, from http://www.dataschool.io/simple-guide-to-confusion-matrix-terminology/