ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Master Thesis – Data Science and Marketing Analytics (DSMA)

Title: Emotion, Relevance, and Intent – A Text Analytics Investigation Into the Performance of SERP Snippets

Student Name: Shomit Ghosh
Student Number: 557811

First Assessor: Dr Bas Donkers
Second Assessor: Dr Pieter Schoonees

Final Version Date: 13th August 2024

Abstract

This master's thesis aims to investigate the composition of SERP snippets from a text analytics perspective. In particular, the paper looks at how the following characteristics of snippets are associated with snippet performance (measured by the click-through rate, CTR): the emotions that are present within snippets, the textual similarity (relevance) between the search query and the snippet, and the search intent. The study makes use of SEO data from the marketing analytics platform, Similarweb. As such, this paper aims to enrich the body of literature surrounding text analytics for SERP research by using an alternative approach (as opposed to search query logs). Aside from multiple linear regressions, this study also makes use of random forests and global interpretation techniques, such as permutation feature importance and partial dependence plots, to add a layer of robustness to the findings. From a managerial perspective, this paper hopes to provide marketers with practical insights into how to formulate SERP titles and descriptions. For example, in the presence of negative emotions, relevance has a stronger impact on CTR than when a snippet has predominantly positive emotions. Thus, marketers must account for the context of searches when deciding on a SERP snippet title and description. Overall, this paper provides insights into relationships between textual features of SERP snippets and their performance.

1. Introduction

1.1. Overview

Since the advent of the worldwide web, search engines have undergone immense change. Early iterations of search engines offered basic indexation tools and very limited interactivity. In contrast, the search engines of today, among other features, offer hyper-personalization and are incorporating advanced natural language processing (NLP) to make search even more ‘natural’.

With these advances, search engine marketing (SEM), comprising of primarily of search engine optimization (SEO) and search engine advertising (SEA), has also grown tremendously as an industry. In fact, SEA is often regarded as the first big wave of digital advertising, preceding social media and retail media (Feger, 2023), and is an industry that has shown consistent, sustained growth. According to Statista (2023), the projected spending on search advertising in 2024 is a staggering \$306.5Bn (a 9.7% increase from 2023) – for scale, that is \$6Bn more than the 2022 GDP of Romania (The World Bank, 2022). Furthermore, a recent HubSpot (2024) survey of more than 1,000 marketing professionals found that SEO was considered one of the most important marketing channels, after short-form video content and influencer marketing.

The SEM ecosystem is incredibly large and complex, with a myriad of moving parts. Moreover, given the scale of the industry and its impact on modern commerce and information retrieval processes, it is understandable that SEM receives a lot of academic and managerial attention. In particular, topics (among others) such as sponsored search, auctioning systems, and SEO are heavily studied. The focus of this research paper, however, is to investigate a relatively understudied area of SEM – one that incorporates text analytics in the investigation of the effectiveness of search engine results page (SERP) snippets (see Figure 1 for an example of an individual snippet on a SERP). By approaching the analysis with a text analytics perspective, we aim to extract practical insights into the composition of SERP snippets, aiding marketers in their formulations of SEO features such as titles and meta descriptions of webpages.

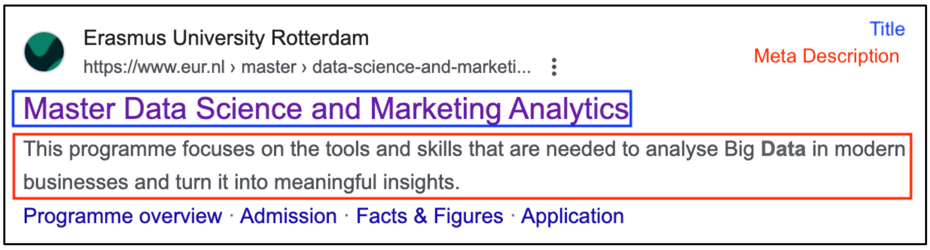


Figure 1. An example of a SERP snippet (Blue: snippet title, Red: meta description)

1.2. Problem Statement

The central focus of this research is the relationship between textual features of SERP snippets and snippets' effectiveness. For the purposes of this study, snippet effectiveness, the dependent variable (DV), is given by a proxy of their click-through rate (CTR), providing insights into the ratio between the number of searches for a given query and the resulting clicks to a given website. As for textual features, we will consider emotion and relevance (a more detailed discussion on these will be provided in the next chapter). Combining all of this, we arrive at the central research question for this study, after which we will proceed to discuss the motivations for pursuing this research:

To what extent do textual features influence the effectiveness of SERP snippets?

1.3. Academic Relevance

The existing literature surrounding the use of text analytics for SERP research is rather limited, especially in a marketing context. As such, this study borrows from studies in the fields of information retrieval, psychology, and information technology (discussed in greater detail in the next chapter). Therefore, one of the key motivations for this study is to enrich the body of literature that looks at the relationship between textual features and SERP snippet effectiveness. For example, one of the focal papers that investigates "the emotion profile of web search" (Kazai, Thomas, & Craswell, 2019) suggests incorporating more lexicons and user intent for further research – something this study implements. Overall, the academic relevance of this paper is that it aims to shed more light on the role of textual features on online search processes. Given the importance of information search in customer journeys, this study is indeed of relevance in the marketing space.

1.4. Managerial Relevance

As mentioned earlier, SEM is a field of digital marketing that has seen incredible growth over the years and consistently receives large amounts of investment each year. Given the scale of the SEM operations, it is in the best interests of marketing professionals to continually strengthen their understanding of how consumers interact with SERP snippets. For example, more than 94% of users only focus on the first page of SERPs – moreover, these users also tend to change their search keywords instead of looking at the next page, when they are not satisfied with the results (Sharma, Shukla, Giri, & Kumar, 2019). Additionally, approximately 63% of users only consider the top three SERP listings (Sharma, Shukla, Giri, & Kumar, 2019). Of course, while SEA plays a big role in the placement of SERP snippets, it is also worth investigating ways to optimise the snippets themselves, by tailoring the titles and meta descriptions. As such, by employing a text analysis approach, we aim to uncover insights that can be applied directly by practitioners in their decisions when it comes to the composition of the title and meta description tags that are to be displayed in the SERP snippets.

Overall, the managerial relevance of this study is that it aims to aid marketing professionals to better understand how they can optimise SERP listings to maximize CTR.

1.5. Thesis Outline

The next section of this paper will provide a theoretical framework, consisting of a literature review that will form the basis of the hypothesis testing and conceptual framework. Then, more insights will be provided on the data that was analysed, followed by a chapter on the methodologies that were employed for this research. Finally, the thesis will conclude with chapters on the results and conclusions, respectively – this is where the main insights from the research will be revealed, while offering an evaluation of the study and suggestions further research.

2. Theoretical Framework

2.1. Overview

This section of the paper outlines the theoretical foundations upon which this study is built. In particular, the framework consists of a literature review (see Appendix A for a summary of the literature in tabular format) in which related literature is synthesized for the purposes of hypothesis development. Following this, a visual representation of the hypotheses is provided in the conceptual framework, outlining the relationships between various textual features, controls, and the DV.

2.2. Literature Review

The central research question aims to study the relationships between textual features and SERP snippet effectiveness. The textual features in question relate to emotion and relevance. For the purposes of this study, we will focus on the emotion present in the text (snippet meta description), rather than the emotions/moods experienced by the users of the search engines. Whereas relevance looks at how useful SERP snippets are based on corresponding search queries. Moreover, aside from the key areas of focus, the literature considered also looks at the variables that need to be controlled for. The subsequent sections borrow insights from various fields (including marketing, information retrieval, and information technology) for hypothesis development.

2.2.1. Emotion

To contextualise this section of the literature review, we turn to a paper that formulates a "model of emotions and mood in the online information search process" by Lopatovska (2014). The latter part, mood which is more long-lasting and "does not have a clear ending or beginning (Lopatovska, 2014), is not relevant for this study – however, this paper provides foundations that indicate that emotions play a vital role in the information search process. For example, the most common facial expressions that resulted during the search processes were surprise and neutral. The author also outlines that "emotions are known to play an integral part of information search processes as they affect a searcher's attention, memory, performance, and judgments" (Lopatovska, 2014), thus motivating this part of the study.

As mentioned in Chapter 1, one of the focal papers for this study is the one that investigates the "emotion profile of web search" (Kazai, Thomas, & Craswell, 2019). Using query logs obtained from Bing over a fourth-month period in 2019, the researchers looked at the emotion profile of search results using a lexicon approach. The lexicons in question were SentiWordNet (positive, negative, objective) and EmoLexData (afraid, amused, angry, annoyed, don't care, happy, inspired, and sad emotions). Furthermore, the paper looks at whether differences in emotion profiles of search results exist between clicked and non-clicked results, relevant versus irrelevant results, page rank, and whether the search queries related to controversial or mundane topics. Notably, they found that clicked results were "significantly more positive and happy than not-clicked results" (Kazai, Thomas, & Craswell, 2019) – this is particularly interesting because they also found that results ranked higher tended to be less emotionally charged – the authors explained this effect by suggesting that the top-ranked results were often navigational in nature and, therefore, were relatively "emotionless". Thus, it seems that even if emotional results rank lower, they are more frequently clicked by users compared to emotionless results, suggesting a potentially stronger influence of emotion on search behaviour compared to SERP rank (Kazai, Thomas, & Craswell, 2019). Moreover, the logistic model designed to model the propensity click found that a "purely positive document title and snippet" was associated with 6.64 points higher log odds of a click, compared to an emotionless counterpart (Kazai, Thomas, & Craswell, 2019). Finally, the authors suggest incorporating user intentions (discussed further in section 2.2.4) and more lexicons. Therefore, building on the research by Kazai et al. (2019), this study employs the NRC Word-Emotion Association lexicon (also known as EmoLex) which captures the following emotions: anger, anticipation, disgust, fear, joy, sadness, surprise, and trust, and sentiments: positive/negative (Mohammad & Turney, 2011).

The next relevant paper builds on the study by Kazai et al. (2019) and applies it to the context of online search in classrooms (Landoni, Pera, Murgia, & Huibers, 2020). The motivation for this study is that the researchers wanted to diversify the approach to researching information retrieval processes away from just "relevance, readability, and reliability of retrieved documents" (Landoni et al., 2020). As with the previous study, the research by Landoni et al. (2020) looks at using a lexicon approach to analysing search query logs gathered from classrooms (third to fifth grade) in the US, Italy, and Switzerland between November 2018 and July 2019. The study also made use of direct observations by teachers in the classroom. Interestingly, some of their findings do not align with the ones discussed in the previous paper by Kazai et al. (2019). For example, the authors found no significant differences in the emotion profiles of search results across page rank positions. Furthermore, they also found no significant differences in the emotion profiles of clicked versus non-clicked results (Landoni, Pera, Murgia, & Huibers, 2020), putting in question the relevance of emotion in the effectiveness of SERP snippets.

Finally, the paper by Milton and Pera (2020) also builds on the work of Kazai et al. (2019) by employing a similar methodology but to investigate how SERP snippets impact users suffering from mental health disorders (MHDs). The authors posit that search engines are considered to be a "persuasive technology" – one that can "change the behaviours or attitudes of individuals" (Milton & Pera, 2020), arguing that search engines should be further optimised for users suffering from MHDs, as they can have varying reactions to persuasive technology. Their methodology consisted of developing synthetic search query logs for traditional users and those with MHDs, where MHD queries were obtained from online forums commonly frequented by people with MHDs. Using these queries, they extracted the resulting SERP snippets from the Google API. While the authors agree with Kazai et al. (2019) that most results were objective in nature, they found that MHD-associated search queries resulted in "more polar sentiments and negatively charged emotions" (Milton & Pera, 2020). While the outcomes of the study are outside the scope of this study, they further reinforce the idea that there indeed exists a relationship between users of search engines and the emotions associated with SERP snippets.

Considering the relevant literature, it seems that there exists some variance with regards to the impact of emotions in SERP snippets on SERP snippet effectiveness. Therefore, to further test the relationship, the following emotion-related hypotheses will be tested, where anticipation, joy, surprise, and trust are considered positive emotions, whereas anger, disgust, fear, and sadness are considered negative emotions:

H1a: Snippets that possess higher levels of anticipation are associated with higher CTRs.

H1b: Snippets that possess higher levels of joy are associated with higher CTRs.

H1c: Snippets that possess higher levels of surprise are associated with higher CTRs.

H1d: Snippets that possess higher levels of trust are associated with higher CTRs.

H1e: Snippets that possess higher levels of anger are associated with lower CTRs.

H1f: Snippets that possess higher levels of disgust are associated with lower CTRs.

H1g: Snippets that possess higher levels of fear are associated with lower CTRs.

H1h: Snippets that possess higher levels of sadness are associated with lower CTRs.

2.2.2. Relevance

Given that primary function of search engines is to provide users with information that corresponds to their search queries, relevance of SERP snippets is expected to be a key driving factor of CTR. For this section, we consider studies in information retrieval and marketing.

The first paper, from the field of advertising, by Kononova et al. (2020) considers the effects of displaying relevant and irrelevant ads on various aspects of users' brand recognition, click intention, etc. While the study operates in a social media context, whereas the focus of this thesis is on SERPs, we argue that the contexts are actually quite similar. Kononova et al. (2020) focus on ads that are incorporated in a set of organic online stories. Similarly, SERPs are displayed as a mixture of organic and sponsored snippets. Therefore, we believe that we can attempt to extrapolate and test the concept of relevance in the SERP space. The findings of the study are in line with expectations – irrelevant ads were less likely to be clicked on, compared to relevant ads (Kononova, Kim, Joo, & Lynch, 2020). Notably, click intention was derived by asking respondents to evaluate how likely they were to click on ads (likely/not likely), rather than observed clicks (Kononova, Kim, Joo, & Lynch, 2020). Therefore, this thesis aims to build on these findings by using (estimated) observed click data in the form of CTRs.

The paper discussed in section 2.2.1 by Kazai et al. (2019) also looked at relevance of SERP snippets (alongside its interaction with emotion profiles). The authors used relevance labels from the query logs to label query-URL pairs as relevant or irrelevant. As expected, the observed click rates for relevant documents were higher than those for irrelevant documents (Kazai, Thomas, & Craswell, 2019). Also, the findings show that there exists an interaction effect between relevance and emotion profiles – in particular, they find that for "relevant results the effect of sentiment is greatly exaggerated compared to irrelevant results" (Kazai, Thomas, & Craswell, 2019).

Overall, the limited literature shows that indeed relevance has an impact on click behaviour. Moreover, there appear to be additional interaction effects that will also be investigated further – one with emotion and one with search intent (see section 2.2.4). Thus, the following hypotheses can be developed from this section:

H2: More relevant snippets are associated with higher CTRs.

H2a: The effect of relevance on CTRs is (positively/negatively) moderated by the snippet sentiment (positive/negative).

2.2.3. Search Intent

Search intent is often broadly categorized as one of the following: informational, navigational, and transactional (Broder, 2002). These are also the definitions used by Similarweb, the primary data source for this thesis (more on this in Chapter 3). Informational search intent refers to the scenario where users are looking to "learn something or get information" (Similarweb, n.d.). Navigational intent occurs when a user enters a search query with goal of finding a particular domain or webpage. Finally, transactional search intent is associated with making transactions (e.g., booking hotels, etc.).

The review of eye-tracking studies in the context of SERPs by Lewandowski and Kammerer (2021) found no systematic differences in SERP-viewing behaviour across the three search intents in the 41 papers that were considered in the review. This contrasts to the findings of the paper that looked at "construal matching in online search" (Humphreys, Isaac, & Wang, 2021). Humphreys et al. (2021) look at the relationship between users' mindsets and the search queries they generate, and in turn how likely they are to click on SERP results, along different points in the customer journey (informational versus transactional). Indeed, they find that “that when consumer mindsets are more abstract (more concrete), consumers generate textual search queries that use more abstract (more concrete) language” (Humphreys, Isaac, & Wang, 2021). They also find that SERP results that match their mindsets (i.e., search intents) are more likely to be clicked (Humphreys, Isaac, & Wang, 2021). Given that earlier stages of the customer journey have informational needs and later stages have transactional needs, and given the inherent differences in the two stages, we should expect differences in click behaviour across the search intents.

The paper by Shi and Trusov (2021) found that search intent is indeed one of the key predictive factors when it comes to users’ inspection process of SERPs and their SERP snippet click behaviour – in particular, they found that users with transactional (navigational) intent had the highest (lowest) likelihood of “going below the fold” (Shi & Trusov, 2021), meaning that they were more likely to scroll past the initial set of SERP snippets.

Furthermore, the paper by Nagpal and Peterson (2021) found the relationship between relevance and click behaviour in SERPs to be a bit more nuanced. They found that content relevance was only a driving factor of clicks when the user was further along the customer journey (indicated by a transactional search intent, versus an informational search intent).

Considering the views presented in this section and the previous section around relevance, and given Kazai et al.’s (2019) recommendation to incorporate search intents into the study of emotional profiles of web search, the following hypotheses be developed:

H3: Transactional searches are associated with the highest CTRs, and informational searches are associated with the lowest CTRs.

H3a: The effect of relevance on CTRs is moderated by search intent – with transactional intent having the strongest influence and informational intent having the weakest influence, and navigational intent falling in between the two.

2.2.4. Controls: SERP Features and Length

Finally, this part of the review will consist of a brief discussion of the control variables for this study – namely, SERP features and, snippet title and meta description lengths. Although not part of the conceptual framework, it is a worthwhile exercise to explore SERP characters that can influence click behaviour but are outside the scope of our investigation.

SERP features, found in most modern search engines, go beyond displaying the most basic snippets (consisting of just a title and meta description). Figure 2 provides some examples of these features. As far as its influence on click behaviour is concerned, it can have varying impacts, depending on the search task. For example, if a user wanted to find out the address of Erasmus University Rotterdam, they do not need to engage in any clicks in the presence of the knowledge card SERP feature (see Figure 1). Similarly, if a user is looking for exchange rates, modern search engines can natively display such information in SERP features, thus leading to fewer clicks required by users. SERP features form “rich snippets” (Marcos, Gavin, & Arapakis, 2015) that have shown to influence users’ viewing behaviour of SERP snippets, in the form of enhanced attention capture.

As for snippet lengths, the length of snippet titles will be controlled for. Furthermore, a maximum of 50 words will be analysed from the meta descriptions to account for the fact that most search engines do not display more text than that, as allowing more text than that to be displayed would hamper the user experience. A study by Cutrell and Guan (2007) found that there were no main effects of snippet length on users’ viewing behaviour, but there were significant interactions with search task types (intent). Given that snippet lengths are not the focus of this study, we simply control for snippet length (alongside SERP features).

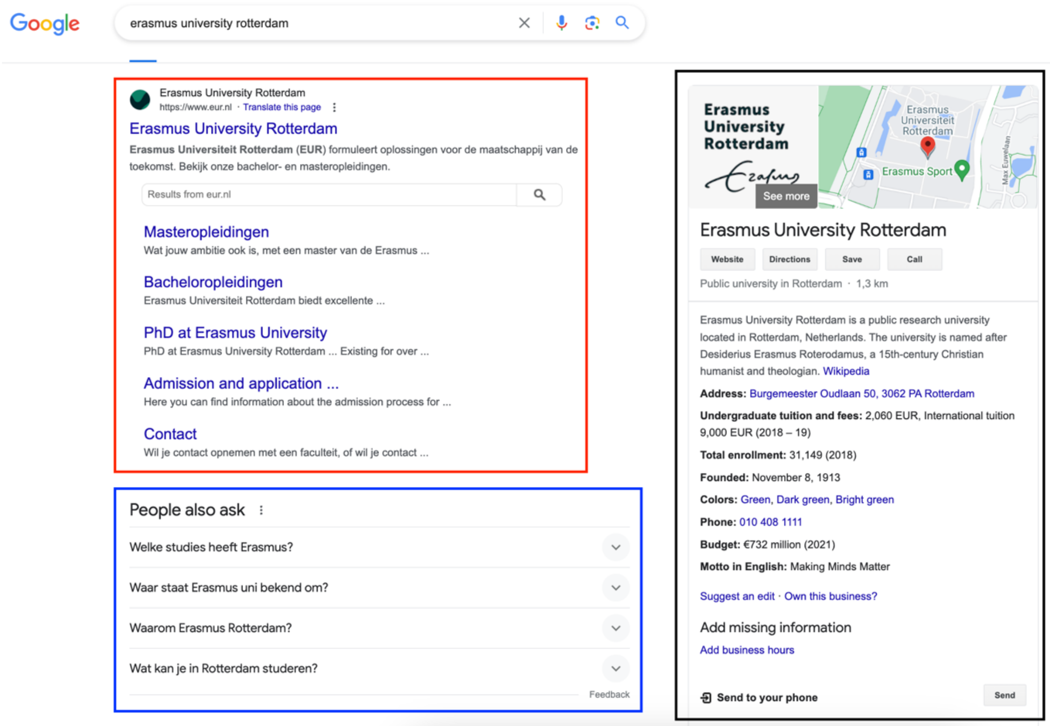


Figure 2. Examples of SERP features. Red: expanded links, Blue: related questions, Black: knowledge card.

2.3. Conceptual Framework

Provided below is a visual representation of the conceptual framework that is being explored in this study, with the key areas of focus being the emotion profile of SERP snippets, relevance, and search intent.

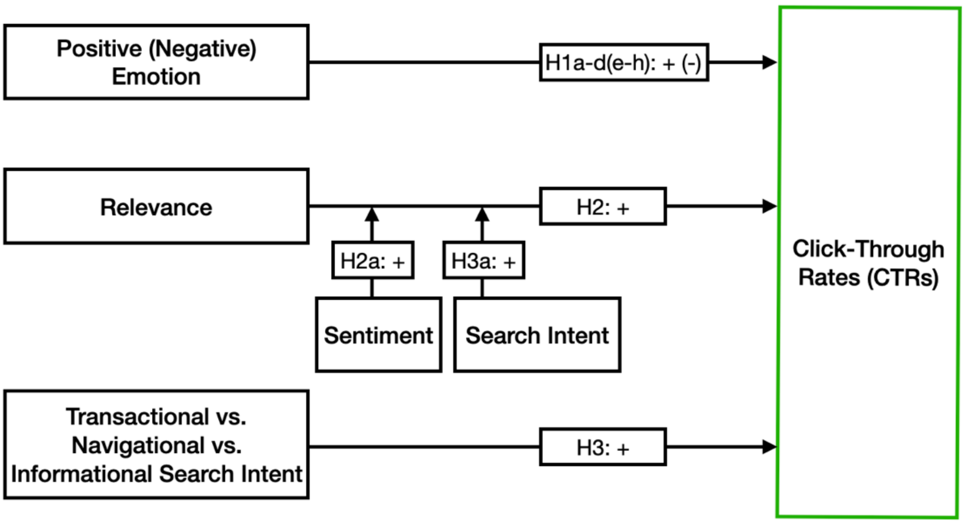


Figure 3. A visual representation of the overall conceptual framework.

2.4. Summary of Hypotheses

Overall, this theoretical framework provided relevant literature that formed the basis for hypothesis development. Given the differences in findings and gaps in research, this thesis aims on adding to the growing body of literature surrounding SERP snippets and textual analytics. Provided in Table 1 is a summary of the hypotheses being tested in a tabular format.

Table 1. Summary of hypotheses

H	Hypothesis
1a	Snippets that possess higher levels of anticipation are associated with higher CTRs.
1b	Snippets that possess higher levels of joy are associated with higher CTRs.
1c	Snippets that possess higher levels of surprise are associated with higher CTRs.
1d	Snippets that possess higher levels of trust are associated with higher CTRs.
1e	Snippets that possess higher levels of anger are associated with lower CTRs.
1f	Snippets that possess higher levels of disgust are associated with lower CTRs.
1g	Snippets that possess higher levels of fear are associated with lower CTRs.
1h	Snippets that possess higher levels of sadness are associated with lower CTRs.
2	More relevant snippets are associated with higher CTRs.
2a	The effect of relevance on CTRs is (positively/negatively) moderated by the snippet sentiment (positive/negative).
3	Transactional searches are associated with the highest CTRs, and informational searches are associated with the lowest CTRs.
3a	The effect of relevance on CTRs is moderated by search intent – with transactional intent having the strongest influence and informational intent having the weakest influence, and navigational intent falling in between the two.

3. Data and Methodology

3.1. Data Collection

The data collection process for this study consisted of two stages. The first stage involved collecting search data from the data aggregation service, Similarweb, that focuses on gathering marketing analytics data. In particular, data was gathered from their ‘Search 3.0’ dataset (Similarweb, 2024) on the organic search performance of 30 brands across six industries (e-commerce, finance, food and drink, health, music, and travel). The brands were selected such that they were the top performers in their industries and primarily operated in English, for maximum opportunities for clean, usable data. For each brand, the dataset provided data on the performance and associated characteristics of the 1,000 most popular search keywords (by clicks). By using data on keywords rather than query logs, we effectively remove the impact of position ranking on CTR – focusing only on the ‘top URLs’ associated with each of the search keywords. The top URL is simply the URL that is estimated (by Similarweb) to be driving the most traffic to the focal website, for a given keyword (regardless of the position of the snippets). By doing so, we can direct the focus of the study further towards the textual features, rather than position, etc. Hence, this is also one of the ways this thesis builds on the literature provided in the previous chapter, by exploring alternative methods (to query logs) of data collection. Overall, the sample of URLs consists of the URLs associated with the top keywords, they are not necessarily the top-performing URLs by themselves overall – this allows us to eliminate the impact of position on CTRs.

The second stage of the data collection process consisted of parsing text data on snippet titles and meta descriptions, stored in the ‘head’ section of HTML documents (see Figure 4 for an example). This was done using the *rvest* package in R (Wickham, 2024). In particular, the ‘top URL’ associated with each keyword was fed into an R function that scraped the contents of the “title”, “description”, “og:title”, and “og:description” meta tags for each URL. The “og” or Open Graph titles and descriptions were also extracted as a form of redundancy measure – for example, we found that in certain cases the og tags were filled but not the default ones, in which case we copied over the content from those tags. After preliminary cleaning, the raw dataset consisted of 5,888 observations.

Figure 4. An example showing how snippet titles (red) and descriptions (green) are stored in the head section of HTML documents. The source code displayed was pulled from the website linked to the snippet title displayed.

3.2. Data Wrangling

There were several transformations performed on the raw data to end up at the final dataset (Table 2 provides a summary and descriptions of the final set of variables that were considered for analysis).

Firstly, the dependent variable, CTR, was constructed by dividing the number of clicks by the number of searches associated with each keyword. Here, it is worth noting that the data provided by Similarweb does not always consists of raw data – instead, it is often a mixture of data from “direct measurement”, Similarweb’s “contributory network”, and their own data models (Similarweb, 2024) – all of this provides an aggregation of search data. Therefore, although CTR is usually considered a direct measure, in this study it should really be considered more of a proxy of the direct CTR (given that the CTR being considered in this study is developed using estimates of clicks and searches).

After extracting text data, it was found that several websites did not assign title and meta descriptions at all. In these cases, the observations were removed (after checking for significant differences in CTR, of which there were none).

Next, Similarweb provides search intents in the form of a list. In certain cases, Similarweb deemed certain keywords to be associated with multiple intents, with the first item in the list representing the dominant search intent (Similarweb, 2024). Accordingly, only the dominant search intents were stored for further analysis.

Moving to text data, the snippet titles and meta descriptions were pre-processed in accordance with standard practice for sentiment analysis – this involved, for example, removing special characters, extra blank spaces, removing numbers, etc. This text data was then analysed using the *syuzhet* package in R (Jockers, 2023), where each meta description was assigned scores on the eight emotions being considered (more on the methodology in the next section).

Finally, the relevance measure, given by the Jaccard coefficient, was also a calculated measure. However, a more detailed description of this is provided in the next section around the employed methodologies.

Table 2. Summary of all variables being considered for this study.

Variable	Type	Description
CTR	Numeric	Expressed as a percentage, CTR looks at the ratio between the number of clicks and total search queries associated with a given keyword.
Search Intent	Factor	The dominant search intent associated with each keyword, taking a value of either informational, navigational, or transactional.
SERP Features	Factor	Eight individual variables that look at which SERP features (news, video, apps, images, related questions, knowledge cards, local information, and expanded links) are associated with the given keyword. For each of the features, the variable can take a value of TRUE or FALSE.
Anger	Numeric	For each word in the meta description, the NRC Word-Emotion Association Lexicon assigns a score of 1 or 0 (associated with anger or not associated with anger, respectively). The scores are then summed for each meta description.
Anticipation	Numeric	Similar as above but with anticipation.
Disgust	Numeric	Similar as above but with disgust.
Fear	Numeric	Similar as above but with fear.
Joy	Numeric	Similar as above but with joy.
Sadness	Numeric	Similar as above but with sadness.
Surprise	Numeric	Similar as above but with surprise.
Trust	Numeric	Similar as above but with trust.

Negative	Numeric	Employs a similar methodology as emotions, but here the words are labelled as either negative (1) or not negative (0).
Positive	Numeric	Similar as above, but with positive.
Relevance	Numeric	Given by the Jaccard similarity (coefficient) between the search query and snippet title.
Title Length	Numeric	The number of words used in the title.
Meta Length	Numeric	The number of words used in the meta description.

3.3. Methodology

3.3.1. Emotions and Relevance

Hypotheses 1 and 2a deal with emotions and sentiments, respectively. To extract information on the emotion and sentiment levels, this study employs the NRC Word-Emotion Association lexicon implemented using the *syuzhet* package. For each word, the lexicon has an associated set of emotions and sentiments. For example, according to the lexicon, the word "achieve" is associated with the emotions of joy and trust and has a positive sentiment (Mohammad & Turney, 2011) – in this case a score of 1 would be assigned for joy, trust, and positive if our document only consisted of the word "achieve". Thus, for each meta description, the emotion and sentiment scores are made up of the number of words that are associated with the given emotions and sentiments. Of course, a longer meta description is more likely to have a higher score for 'joy', for example, which is why this study controls for the length of meta descriptions.

As for relevance, it is defined as the "degree of overlap or semantic and textual similarity between the webpage content and the search query" (Nagpal & Peterson, 2021). In particular, this study makes use of the Jaccard coefficient – this is given by the ratio between the size of the intersection and the size of the union between the search query and snippet title (Gupta, Saini, & Saxena, 2013). The resulting coefficient can range between zero and one, where zero implies that there is absolutely no overlap between the search query and the snippet title. Whereas a value of one indicates that the search query and snippet title consist of identical words (intersection is equal to the union).

3.3.2. Linear Regressions

The bulk of the analysis performed in this study will be done using multiple linear regression models, assuming a linear relationship between the predictor variables and the target variables. In the most basic sense, a multiple linear regression model takes the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \quad (1) \qquad \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \quad (2)$$

Equation 1 consists of the target variable Y, predictor variables X_j, the error term ϵ, and the parameter coefficients given by β_j – it is these coefficients that are estimated using the ordinary least square (OLS) regressor, and these coefficients allow for the prediction of Y, given by $\hat{Y}$ in equation 2. The OLS method aims to optimize the loss function (given by the sum of squared residuals, see equation 3) such that the residuals are minimized (James, Witten, Hastie, & Tibshirani, 2023).

$$RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (3)$$

Aside from the models that focus on the main and interaction effects (hypotheses 1-3), we compute two additional models – one with just the control variables and one with all variables included. The reason for doing this is purely to form baselines, allowing for a comparison of model performance. Therefore, these two models will not be discussed in great detail.

For the purposes of statistical inference, four key assumptions need to hold (Stock & Watson, 2020). First, given the values of the predictor variables, the residuals should be distributed in a way that the mean is equal to zero – this will be checked using a residual plot (residual values on fitted values). Secondly, the residuals must be homoscedastic (equal variance throughout the range of predictor values) – this will be checked using a scale-location plot (Kassambara, 2018). Third, the residuals must follow a normal distribution – this will be checked by observing the Q-Q plot which plots the distribution of the residuals against a theoretical normal distribution. Finally, the observations should be independent of each other – this one is harder to check for. However, to the best of our knowledge, while individual keywords can be similar to each other, they should not be impacted by each other. The data collection process is such that it simply provides data on the 1000 most popular keywords, we have no information on the user journey, etc. Therefore, we will proceed with the assumption that indeed the observations are independent of each other.

3.3.3. Random Forests and Global Interpretations

While most of this study focuses on statistical inference, we also engage in a prediction exercise – this is because one of the managerial goals of this study, aside from providing insights, is to lay some foundations for a potential platform in the future that could allow marketers to test various snippet titles and meta descriptions. For such a platform to exist, we require a robust predictive model – in this study, we employ the ensemble machine learning technique, random forest. Simultaneously, the findings from this prediction exercise will also be used to corroborate the findings derived from the linear regression models.

Random forests build bagged decision trees for regression. Decision trees for regression work by "stratify[ing] the feature space" (James, Witten, Hastie, & Tibshirani, 2023) – this means that observations are split by certain thresholds, creating regions within the predictor space. Thus, to make a prediction, the tree determines which region the predictors fall into and then the output is the mean of that region (James, Witten, Hastie, & Tibshirani, 2023). While decision trees are relatively interpretable, their primary drawback is that they are rather sensitive to changes in data/unseen data (James, Witten, Hastie, & Tibshirani, 2023). Bootstrap aggregation (or bagging) aims to mitigate this problem by running the decision trees on several (hundreds or thousands) of bootstrapped samples, and the final prediction is a mean of all the individual predictions. Random forests go a step further, by "decorrelating" (James, Witten, Hastie, & Tibshirani, 2023) the trees – this is done by selecting a random subset of variables as the predictors for each tree. To optimize the random forest, we will be tuning the hyperparameter (*mtry)* that dictates the size of the subset of variables considered at each tree – this will be done using three-fold cross-validation.

It is important to note that for, both, linear regression models and the random forest, the models are developed using a training set (70% of the full dataset, 2108 observations). The models are then assessed by their predictions made using the test set (20% of the full dataset, 900 observations). The primary measure for model performance used in this study is the root-mean-squared-error (RMSE), providing insights into the difference between predicted and actual values.

Finally, aside from using the random forest for predictive purposes, we will also employ global interpretations technique (permutation feature importances and partial dependence plots) to corroborate our findings from the linear regression models (Molnar, 2022), by making our black-box model more interpretable. Permutation feature importance plots highlight feature importance by looking at the increase in model error when certain variables are omitted from the random forest. For example, if the model error (given by the RMSE) is relatively unchanged when variable A is omitted from the forest, the algorithm considers it a relatively ‘unimportant’ feature (Molnar, 2022). The feature importance plot will be used to check if the features that were considered in the linear regression models are indeed ‘important’. On the other hands partial dependence plots (PDPs) provide insights into the marginal effects of individual features on the random forest’s predictions (Molnar, 2022). The PDPs will be used to confirm the type of relationship (positive/negative/other) between the predictors and the outcome variable, CTR.

3.4. Descriptive Statistics

The tables below provide a tabular summary of the variables being considered for this analysis (see Table 2 for continuous variables and Table 3 for categorical variables). For additional insights, refer to the visuals provided in Appendix A. Table 3 has two striking features – the mean and maximum CTRs are rather high. However, here it is worth noting that the CTR is calculated using estimates (provided by Similarweb) of the number of clicks and search queries, rather than an exact measure from query logs. Furthermore, in the context of SERP, it is plausible that users search for a keyword and come back to the SERP and click on links multiple times, thus explaining the high CTR levels.

Table 3. Descriptive statistics for continuous variables

N = 3008				
Variable	**Min.**	**Max.**	**Mean**	**S.D.**
CTR	50.12	204.57	85.53	12.624
Anger	0	5	0.203	0.4792
Anticipation	0	5	0.764	0.8476
Disgust	0	4	0.134	0.4322
Fear	0	5	0.292	0.6452
Joy	0	6	0.554	0.7668
Sadness	0	4	0.291	0.6014
Surprise	0	3	0.205	0.4637
Trust	0	9	0.995	1.1363
Negative	0	7	0.541	0.9033
Positive	0	11	1.662	1.2627
Relevance	0	1	0.264	0.2256
Title Length	1	37	7.76	3.429
Meta Length	3	47	23.26	6.091

Table 4. Descriptive statistics for categorical variables

N = 3008					
Variable	**True (%)**	**False (%)**	**Informational (%)**	**Navigational (%)**	**Transactional (%)**
Search Intent			56.0	30.2	13.8
News	33.0	67.0			
Video	61.5	38.5			
Apps	32.8	67.2			
Images	82.2	17.8			
Related Questions	88.5	11.5			
Knowledge Card	18.4	81.6			
Local Information	10.7	89.3			
Expanded Links	48.4	51.6			

4. Results

4.1. Model Parameters

Table 5. Model parameters for hypotheses 1a-h, 2, and 3, and for the control model.

	Control	Emotion	Relevance	Search Intent
(Intercept)	100.035*** (1.2579)	97.810*** (1.2346)	96.086*** (1.3699)	97.578*** (1.2377)
Anger		0.097 (0.6022)		
Anticipation		-0.644** (0.3183)		
Disgust		-1.678** (0.6873)		
Fear		-2.034*** (0.4738)		
Joy		0.356 (0.3787)		
Sadness		1.862*** (0.4782)		
Surprise		-0.664 (0.5778)		
Trust		2.387*** (0.2485)		
Relevance			8.114*** (1.1781)	
Search Intent – Navigational				7.764*** (0.6588)
Search Intent - Transactional				3.143*** (0.6780)
SERP Feature - News	-4.253*** (0.5527)	-3.215*** (0.5435)	-4.306*** (0.5467)	-3.462*** (0.5419)
SERP Feature - Video	-3.792*** (0.5351)	-3.049*** (0.5365)	-4.386*** (0.5362)	-3.458*** (0.5192)
SERP Feature - Apps	6.222*** (0.5271)	4.479*** (0.5309)	6.774*** (0.5275)	4.483*** (0.5321)
SERP Feature - Images	-4.470*** (0.6452)	-3.552*** (0.6283)	-4.239*** (0.6389)	-3.939*** (0.6292)
SERP Feature - Related Questions	-3.120*** (0.7242)	-2.541*** (0.7090)	-3.338*** (0.7170)	-2.333*** (0.7046)
SERP Feature - Knowledge Card	-0.647 (0.6716)	-1.127* (0.6544)	-0.188 (0.6676)	-2.441*** (0.6713)
SERP Feature - Local Information	-6.116*** (0.8149)	-6.235*** (0.7966)	-5.743*** (0.8078)	-6.596*** (0.7905)
SERP Feature - Expanded Links	-0.943* (0.5066)	-0.772 (0.4968)	-0.149 (0.5141)	-3.638*** (0.5465)
Title Length	-0.580*** (0.0693)	-0.619*** (0.0681)	-0.402*** (0.0732)	-0.553*** (0.0673)
Meta Length	-0.043 (0.0388)	-0.070* (0.0388)	-0.037 (0.0384)	-0.036 (0.0376)
N	2108	2108	2108	2108

Note: Standard errors are reported in parentheses; significance stars correspond to the following significance levels: * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 6. Model parameters for hypotheses 2a and 3a.

	Relevance : Sentiment	**Relevance : Search Intent**
(Intercept)	94.163***	97.810***
	(1.4025)	(1.2346)
Relevance	10.181***	13.401***
	(1.7379)	(1.4201)
Positive	2.023***	
	(0.2745)	
Negative	-2.451***	
	(0.4286)	
Relevance : Positive	-2.966***	
	(0.7868)	
Relevance : Negative	6.972***	
	(1.3645)	
Search Intent – Navigational		10.865***
		(0.8403)
Search Intent - Transactional		5.376***
		(1.1414)
Relevance : Search Intent - Navigational		-13.379***
		(2.5090)
Relevance : Search Intent - Transactional		-6.973**
		(3.0448)
SERP Feature - News	-3.421***	-3.469***
	(0.5439)	(0.5308)
SERP Feature - Video	-4.304***	-4.473 ***
	(0.5278)	(0.5195)
SERP Feature - Apps	6.426***	4.846***
	(0.5203)	(0.5288)
SERP Feature - Images	-4.321***	-3.746***
	(0.6247)	(0.6173)
SERP Feature - Related Questions	-3.121***	-2.446***
	(0.7099)	(0.6938)
SERP Feature - Knowledge Card	-0.880	-2.340***
	(0.6603)	(0.6639)
SERP Feature - Local Information	-5.708***	-6.082***
	(0.7906)	(0.7761)
SERP Feature - Expanded Links	-0.250	-2.673***
	(0.5094)	(0.5452)
Title Length	-0.385***	-0.312***
	(0.0717)	(0.0709)
Meta Length	-0.068*	-0.043
	(0.0403)	(0.0369)
N	2108	2108

Note: Standard errors are reported in parentheses; significance stars correspond to the following significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

4.2. Controls (SERP Features, Snippet Title and Meta Lengths)

Before analysing the results of the main and interaction effects, we provide a brief discussion of the outcomes of the control variables – namely, the SERP features, title length, and meta length. Starting with the SERP features, the coefficients are consistently negative across all models, with the exception for ‘Apps’ which consistently has a positive coefficient. Except for ‘Apps’ these, mostly significant, results are very much in line with expectations (as discussed in section 2.2.4) – the presence of SERP features (except for ‘Apps’) is associated with lower CTRs, ceteris paribus. This is likely due to the fact that SERP features aim to enhance user experience by providing intuitive information (in the form of knowledge cards, etc.) relating to users’ search queries to shorten their information search process and, therefore, is inherently designed to reduce the number of clicks. The significant positive coefficients for ‘Apps’ are rather unexpected. According to Similarweb (n.d.), the ‘Apps’ SERP features allows users to interact with certain web apps within the environment of the search engine – hence, it is possible that usage of these lighter versions of apps, encourages user to click further into the SERP snippets – however, this can only be confirmed by further investigation.

Next, we see that title lengths also have consistently negative, significant coefficients across the models with values ranging from -0.619 to -0.312. Therefore, on average, an increase in the title length by one word is associated with a decrease in CTR of between 0.619 and 0.312 percentage points. Similarly, meta lengths are also negatively associated with CTRs, but this relationship is mostly insignificant – this is expected and forms a ‘sanity check’, because (as mentioned in section 2.2.4) the snippet meta descriptions were limited to a maximum of 50 words, thus removing much of the variation with respect to meta lengths.

4.3. Emotion (Hypotheses 1a-h)

Moving onto the model concerning the emotional profiles of searches, we find a mixture of significant and insignificant results. First, let us focus on the ‘positive’ emotions as listed in hypothesis 1a – namely, anticipation, joy, surprise, and trust. In this context, we see two significant results and they are for ‘anticipation’ and ‘trust’. Interestingly, ‘anticipation’ has a negative coefficient with a value of -0.644 whereas ‘trust’ has a positive coefficient of 2.387. The latter aligns with expectations – snippet titles that inspire more trust are associated with higher levels of CTR. This also fits the story painted by Kazai et al. (2019) in which documents with a positive emotional profile are more likely to be clicked on. Moreover, one can argue that anticipation, even though classified as a positive emotion here, relates to less directness – a feature that users searching for information might be more drawn towards.

Looking at the negative emotions (anger, disgust, fear, and sadness), we see three significant results, for disgust, fear, and sadness. All three are negative and thus, aligning with expectations, with fear having the largest magnitude of 2.034. Therefore, on average, adding a word connected with fear in the snippet meta description is associated with a decrease in CTR of 2.034 percentage points, ceteris paribus.

Overall, hypotheses H1d, H1f, H1g, and H1h were supported by this model. Considering the number of hypotheses supported and the magnitudes of the coefficients (when weighted equally), it does appear that negative emotions are associated with stronger drop-offs in CTR as compared to searches with positive emotion profiles which, again, aligns with the study by Kazai et al. (2019). The implications of these findings will be discussed further in the next chapter.

4.4. Search Intent (Hypothesis 3)

As discussed in section 2.2.3, the literature seemed to be relatively divided on the extent to which search intents influence user behaviour and search patterns. The hypotheses were formulated to capture the marketing concept of customer journeys and the way in which Humphreys et al. (2021) suggested that informational search intents were likely to be expressed by users in earlier stages of the customer journey. Whereas transactional search intents were associated with users that were closer to purchases and other forms of transaction (Humphreys, Isaac, & Wang, 2021). However, the results of our model do not quite align with that idea of search intent. The results show that both transactional and navigational searches are associated with significantly higher CTRs than informational searches – however, navigational searches are associated with significantly higher CTRs than transactional searches. On average, a navigational search is associated with a CTR of 7.764 percentage points higher, compared to an informational search, ceteris paribus. Similarly, on average, a transactional search is associated with a CTR of 3.143 percentage points higher than an informational search, ceteris paribus.

Overall, given these significant results, H4 is not supported. However, these findings are still valid and relevant, and will be discussed further in the conclusion and discussion section.

4.5. Relevance, Sentiment, and Search Intent (Hypotheses 2, 2a, and 3a)

Moving onto relevance, given by the Jaccard similarity between the search query and the snippet title, we see a significant and positive relationship between Jaccard similarity and CTR. The model output concerning relevance shows that, on average, snippet titles that are identical to the search queries are associated with a CTR of 8.114 percentage points higher than a snippet title that has zero overlap with the search query, ceteris paribus. Alternatively, on average, a 0.1 unit increase in the Jaccard coefficient is associated with a 0.8114 percentage point increase in CTR, ceteris paribus. Thus, H2 is supported by the model.

Diving further into the interaction effects of sentiment and search intent on CTR, we see some significant results. First, looking at the model 'Relevance:Sentiment' we find that relevance, and positive and negative sentiments all yield reasonable and significant coefficients, with the magnitude of negative sentiment being slightly larger than the positive sentiment (2.451 versus 2.023, respectively). However, the results of the interaction effects are very interesting and unexpected – it seems as though the relationship is exactly the inverse of what was hypothesised. The significant, negative interaction term for relevance and positive sentiment suggests that the impact of relevance is diminished in the presence of higher levels of positive sentiment. Conversely, the significant, positive interaction term for relevance and negative sentiment seems to suggest that the impact of relevance on CTR is emphasised in the presence of higher levels of negative sentiment. As such H3a is not supported by this model. The interesting implications of these results will be discussed in the next chapter.

Finally, to aid the analysis of H3b, we will employ a two-way interaction plot (see Figure 5). However, looking at the main effects, we see a very significant effect of search intent and relevance as discussed in 4.4 and earlier in this section, respectively. Both interaction terms are significant at a 5% level, thus supporting H3b. The interaction term for 'Relevance:Navigational' is -13.379 which is very close in magnitude to the positive coefficient of relevance, 13.401. Therefore, compared to informational search intents, the effect of relevance is reduced to a relatively small amount in the context of a navigational search. A similar relationship holds for transactional searches (compared to informational searches), but the decrease is not as strong. The two-way interaction plot in Figure 5 reflects these interaction effects.

Figure 5. Two-way interaction plot of Relevance and Search Intent. Here, it is visible how the impact of relevance is negligible for Navigational searches (given by the relatively flat line).

4.6. Random Forest (Tuning and Global Interpretations)

In this section, we start to shift our focus slightly towards the predictive exercise, using a random forest. The results of this technique will also be used to corroborate some of our findings from sections 4.2-4.5. Firstly, the hyperparameter *mtry* was tuned to select six variables randomly for each tree (see Figure 6), by minimising the loss (given by RMSE).

Using the tuned tree, we employed the *iml* package (Molnar, 2022) to engage in global interpretations, using permutation feature (Figure 7) importance and PDPs (Appendix B), respectively.

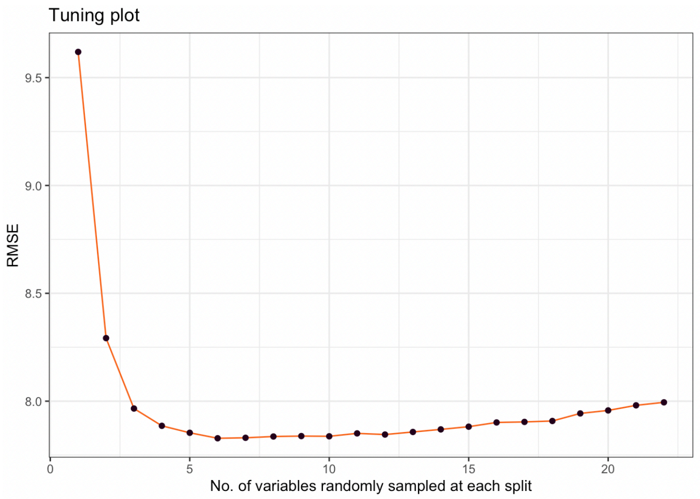


Figure 6. Plot showing model performance during hyperparameter tuning.

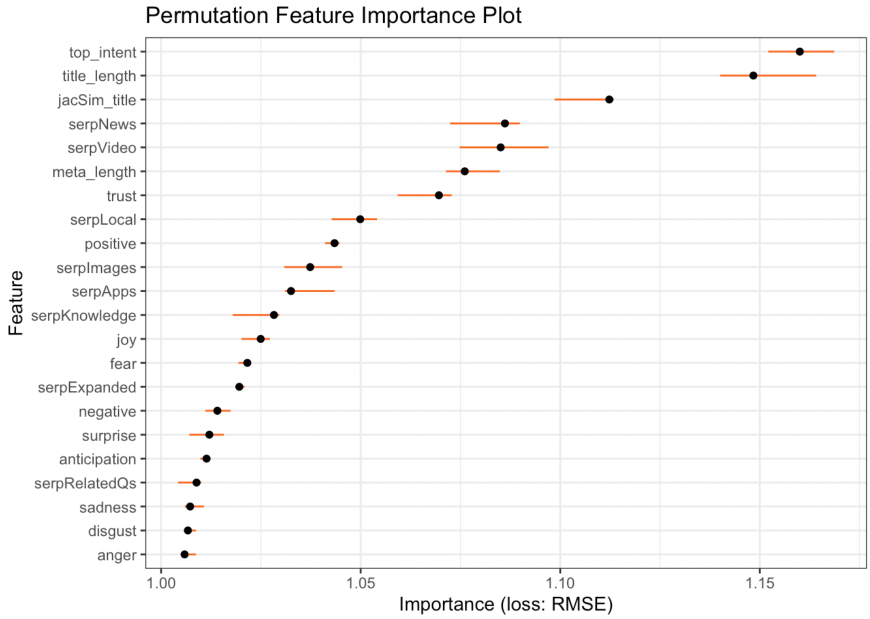


Figure 7. Permutation feature importance plot built using the iml package in R.

In broad terms, the permutation feature importance plot shows us that search intent is the most 'important' feature when it comes to CTRs of SERP snippets. This aligns with our findings that, among the variables of interest, search intents had the largest magnitude of coefficients. Next, relevance is also fairly high up on the list followed by positive emotions ranking in the middle, and negative emotions closer to the bottom. The type of relationship (positive/negative) that was found in the main effects of our study using linear regressions was also confirmed by the PDP (see Appendix B).

Overall, the random forest allowed us to build a more powerful predictive model than the linear regression models (more on this in the next sub-section), while also allowing us to perform a second check on our findings using global interpretations.

4.7. Model Diagnostics

As far as the assumptions of linear regressions are concerned, it appears as though the conditions were satisfied, with a slight risk of heteroskedasticity in some cases. For example, if we look at the model diagnostic plots for the 'Emotion' model (Figure 8, see Appendix C for all model plots), we can see that the 'Residuals vs Fitted' plot is centred around zero and has a fairly horizontal mean. The Q-Q plot also indicates the normality of the distribution of residuals. Finally, the 'Scale-Location' plot suggests a slightly heteroskedastic relationship, as can be seen by the slope around 88 on the horizontal axis.

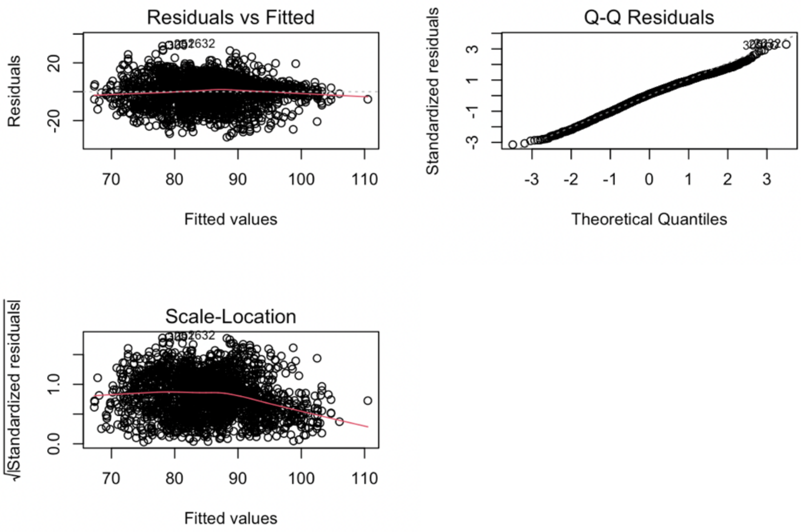


Figure 8. Model diagnostic plots for 'Emotion'.

Next, Figure 9 shows a visual overview of how the models performed, by comparing their training and test RMSEs. The random forest is more than 30% more accurate in its predictions, compared to its linear regression counterparts. As expected, the control model performs the worst but surprisingly, not that much worse indicating that potentially SERP features, and snippet lengths are rather indicative of CTR behaviour. We can also see that the models with interaction terms outperform their main effect counterparts, further validating the fact that the hypotheses being tested were indeed valuable. Finally, we see that the test errors are not massively higher than the training errors, indicating that the models should not be overfitting and relatively robust to new data.

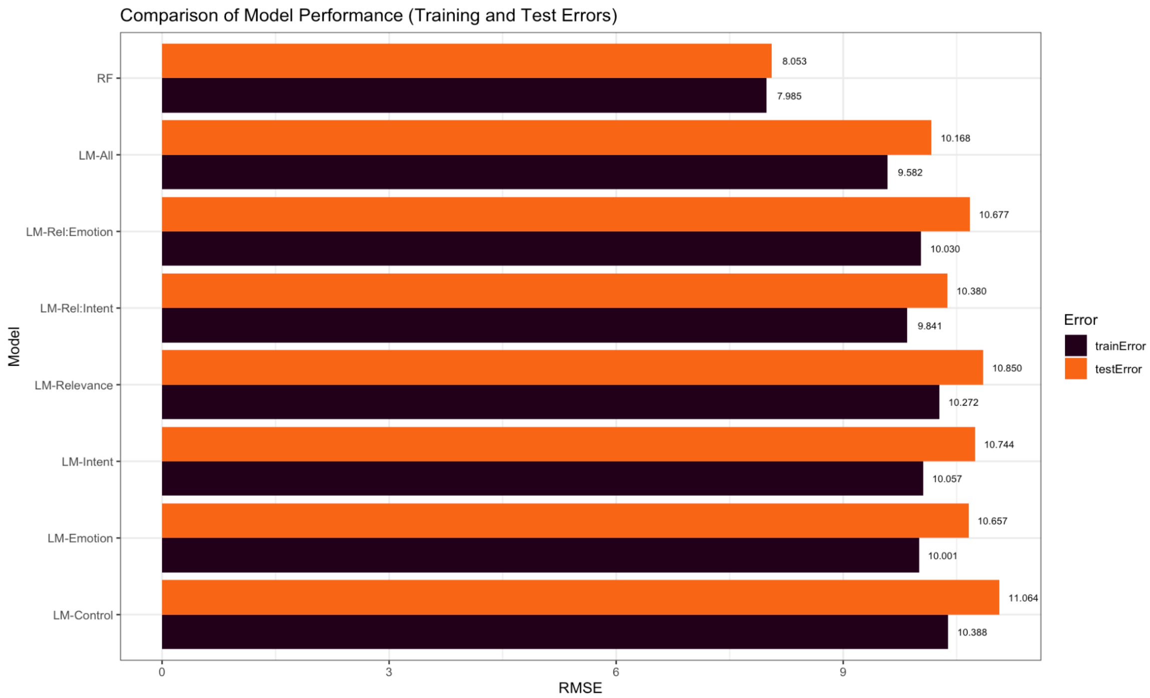


Figure 9. Comparison of model performance using RMSE (RF = Random forest, LM-ALL = Linear regression with all predictors, LM-Rel:Emotion = 'Relevance:Sentiment', LM-Rel:Intent = 'Relevance:Search Intent', LM-Relevance = 'Relevance', LM-Intent = 'Search Intent', LM-Emotion = 'Emotion', LM-Control = 'Control')

5. Conclusion and Discussion

Overall, this study embarked on using text analytics to understand the performance of SERP snippets. In a landscape where large investments are being made annually into SEM, it is crucial for academics and marketing professionals to understand every aspect of the search engine. By focusing on emotion, search intent, and relevance, this study hopes to have added to the growing body of literature around search processes, while offering insights that can be acted on by marketers when constructing SERP snippets.

5.1. Key Findings and Managerial Implications

In terms of emotions, we see that, as expected, negative emotions such as disgust and fear are negatively associated with CTR, and positive emotions such as trust are associated positively with CTR. It is also worth noting that the magnitude of the effect of negative emotions on CTR is slightly larger than that of the positive emotions. While these are expected, when we throw relevance into the mix, the results we saw were quite surprising. We see that the impact of relevance diminishes in the presence of positive emotions and magnifies in the presence of negative emotions in the snippets. One potential reason for this is that if a user's query (here being studied using search keywords) is negative to begin with, they might be in a certain headspace to be looking for specific information with a greater focus – similar to the ideas expressed by Milton and Pera (2020) where they claimed that people suffering from mental health disorders (MHDs) are a) more likely to be making 'negative' queries, and b) perceive and click on SERP snippets differently to those that do not have MHDs (it goes without saying that there are ethical concerns here – one should not use negative language in the hopes of maximising CTRs from a MHD-suffering audience). Whereas positive snippets are also more likely to be more generic and less interactive with relevance. Of course, we are not claiming this as the mechanism for this surprising result in our study, but it does offer some insight into potential avenues for further research. For marketers, these findings are relevant because it highlights the importance of appropriate diction when it comes to the snippet title (relevance) and meta description (emotion). For example, in a search space where the queries can be broad (i.e., less specific), it makes more sense to lean into the use of words that inspire trust to ensure that CTRs are maximised.

Continuing on the relevance findings, we found, as expected, that the main effect is positive and significant. Therefore, it is in the best interest of marketers to ensure that SERP snippet titles align with popular search queries as much as possible. From an academic perspective, these findings challenge those shown in the review of eye-tracking experiment (Lewandowski & Kammerer, 2021), by investigating relevance through means other than lab experiments. Combining relevance and search intent also proved to be very insightful. We saw that the effect of relevance was reduced drastically for navigational searches, compared to informational ones. Also, in general, the main effects were such that navigational searches were associated with the highest CTRs and informational searches with the lowest CTRs. Although some research found the effects of navigational to be in between informational and transactional (Shi & Trusov, 2021), arguing from a perspective of customer journeys, it also makes sense that navigational has the highest CTRs – one possible reason for this is that in the case of navigational searches, users already know where they want to end up, whereas informational searches are more exploratory and transactional searches are likely to experience more friction as they are closer to a transaction. For marketers, this means that understanding the search intents for certain keywords is crucial in determining how specific (i.e., more relevant) they need to make the SERP snippets (with respect to search queries). For example, if a keyword is associated with navigational searches, marketers can reduce the time spent on constructing a hyper-specific snippet title and instead, opt for something more generic.

Overall, this provided grounds for marketers to put more thought into how SERP snippets are constructed while also aiming to enrich the body of literature surrounding text analytics in the context of SERPs. This thesis will conclude by providing a section on the limitations of the study and suggestions for further research.

5.2. Limitations and Suggestions for Further Research

There are four key limitations of this study. First, the risk of endogeneity. It might be the case that on a macro-level, SERP snippets are formed in a way that emotions, search intents, and relevance are all account for. However, we aimed to mitigate this risk by collecting data on SERP performance for various brands across industries. In order to make the study and its findings even more robust, we would recommend an experimental approach to test the relationships – by controlling the levels of emotion, relevance, and search intent, the model would effectively account for endogeneity.

The next limitation and suggestion for further research surrounds the measurement systems. This study depended heavily on Similarweb's data collection and aggregation services. While Similarweb is an established provider of marketing analytics tools and datasets, there are aspects of the data collection and aggregation that are more 'black box' in nature. For example, the number of clicks associated with a search keyword is derived from a combination of direct measurement, data from external sources, and modelling by Similarweb. As such, one avenue of further research could look into way of collecting data directly from focal website as much possible – given the increase in tightening cookie policies, this might prove to be a real challenge. However, it certainly could pave the way for robust research surround all things SEM.

Next, although this study built on existing research by using a different lexicon to that used by Kazai et al. (2019), a lexicon-based approach does still have its limitations. While they are fairly easy to implement and interpret, they do not account for context as it uses a bag-of-words approach. Thus, extending this study with the use of word embeddings, for example, could lead to more externally valid results as it would be better suited to capturing the impact of context in relation to the emotional profiles of searches.

Finally, one of the bigger risks to external validity relates to how snippet meta descriptions are displayed in search engines like Google. Being a relatively advanced search engine, Google takes into account a myriad of factors when displaying the snippet meta, aside from just the meta description provided in the HTML meta tag. For example, Google might crawl various parts of a website and display different information to different users – of course, this is possible because Google has access to vast amounts of data for research, given that it is the market leader of search engines. Often the displayed descriptions are very similar to those stored in the HTML meta tags, but there still is a risk of the mismatch which in turn, can impact the external validity of this study. Therefore, we recommend devising a research design that focuses more on analysis based on observed snippet meta descriptions – thus, the research would have to be more experiment-based or would require more technical investigation into how to extract and aggregate actual, observed meta descriptions.

Overall, while this study shed light on a relatively understudied topic, it also opens a range of topics for further exploration. From a managerial perspective, we believe that understanding SERP snippets and search behaviours from a textual and emotional approach can be valuable in the development of new snippet title and descriptions. To close, studying the relationship between search and emotions with a textual approach is particularly relevant, especially as we are in the midst of a large-language-model boom.

6. Bibliography

Broder, A. (2002). A taxonomy of web search. *SIGIR Forum, 36*(2), 3-10.

Feger, A. (2023, March 3). *Why retail media will be the third and biggest wave of digital advertising* . Retrieved from EMARKETER: https://www.emarketer.com/content/why-retail-media-will-third-biggest-wave-of-digital-advertising

Gupta, Y., Saini, A., & Saxena, A. (2013). A Review on Important Aspects of Information Retrieval. *International Journal of Computer and Information Engineering, 7*(12), 1638-1646.

Humphreys, A., Isaac, M. S., & Wang, R. J.-H. (2021). Construal Matching in Online Search: Applying Text Analysis to Illuminate the Consumer Decision Journey. *Journal of Marketing Research, 58*(6), 1101-1119.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023). *An Introduction to Statistical Leaning with Applications in R.*

Jockers, M. (2023, August 11). *syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text*. Retrieved from CRAN: https://cran.r-project.org/web/packages/syuzhet/index.html

Kassambara, A. (2018, November 3). *Linear Regression Assumptions and Diagnostics in R: Essentials* . Retrieved from Statistical Tools for High-Throughput Data Analysis: http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-assumptions-and-diagnostics-in-r-essentials/

Kazai, G., Thomas, P., & Craswell, N. (2019). The Emotion Profile of Web Search. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1097-1100). Association for Computing Machinery.

Kononova, A., Kim, W., Joo, E., & Lynch, K. (2020). Click, click, ad: the proportion of relevant (vs. irrelevant) ads matters when advertising within paginated online content. *International Journal of Advertising, 39*(7), 1031-1058.

Landoni, M., Pera, M. S., Murgia, E., & Huibers, T. (2020). Inside Out: Exploring the Emotional Side of Search Engines in the Classroom. *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (pp. 136-144). Assocation for Computing Machinery.

Lewandowski, D., & Kammerer, Y. (2021). Factors influencing viewing behaviour on search engine results pages: a review of eye-tracking research. *Behaviour & Information Technology, 40*(14), 1485-1515.

Lopatovska, I. (2014). Toward a model of emotions and mood in the online information search process. *Journal of the American Society for Information Science and Technology, 65*(9), 1775-1793.

Marcos, M.-C., Gavin, F., & Arapakis, I. (2015). Effect of Snippets on User Experience in Web Search. *Proceedings of the XVI International Conference on Human Computer Interaction.* Association for Computing Machinery.

Milton, A., & Pera, M. S. (2020). What Snippets Feel: Depression, Search, and Snippets. *Proceedings from the 1st Joint Conference of the Information Retrieval Communities in Europe, CIRCLE 2020.* Boise State University.

Mohammad, S. M., & Turney, P. (2011, July 10). *NRC Word-Emotion Association Lexicon (aka EmoLex)*. Retrieved from Saif Mohammad: https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

Molnar, C. (2022). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable.* https://christophm.github.io/interpretable-ml-book.

Nagpal, M., & Peterson, J. (2021). Keyword Selection Strategies in Search Engine Optimization: How Relevant is Relevance? *Journal of Retailing, 97*(4), 746-763.

Sharma, D., Shukla, R., Giri, A. K., & Kumar, S. (2019). A Brief Review on Search Engine Optimization. *International Conference on Confluence The Next Generation Information Technology Summit (Confluence) .* Noida: IEEE.

Shi, S. W., & Trusov, M. (2021). The Path to Click: Are You on It? *Marketing Science, 40*(2), 344-365.

Similarweb. (2024, March 11). *Introducing Search 3.0: The World's Most Powerful SEO Dataset*. Retrieved from Similarweb: https://www.similarweb.com/blog/updates/product-updates/search-3-0/

Similarweb. (n.d.). *Keyword Search Intent*. Retrieved from Similarweb: https://support.similarweb.com/hc/en-us/articles/7340850354065-Keyword-Search-Intent

Similarweb. (n.d.). *SERP Features Report* . Retrieved from Similarweb: https://support.similarweb.com/hc/en-us/articles/14689873167901-SERP-Features-Report

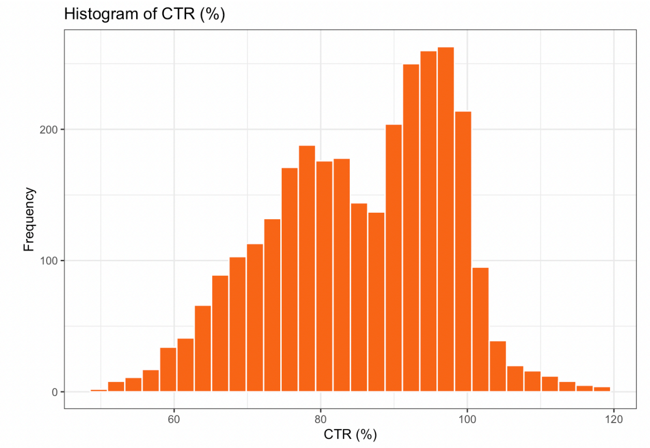
Statista. (2023). *Search Advertising - Worldwide* . Retrieved from Statista: https://www.statista.com/outlook/dmo/digital-advertising/search-advertising/worldwide

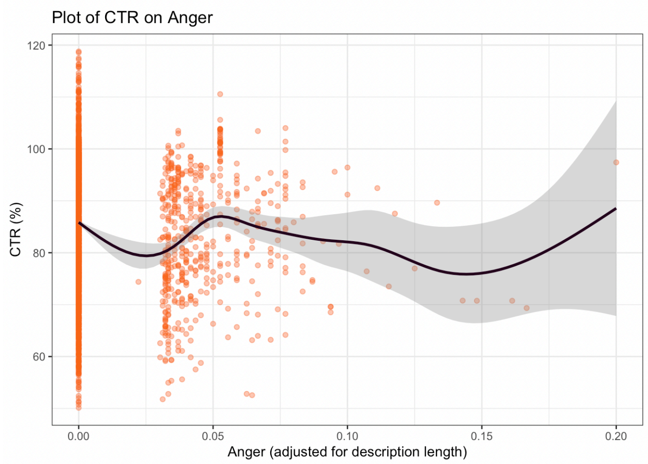
Stock, J. H., & Watson, M. W. (2020). *Introduction to Econometrics.* New York: Pearson.

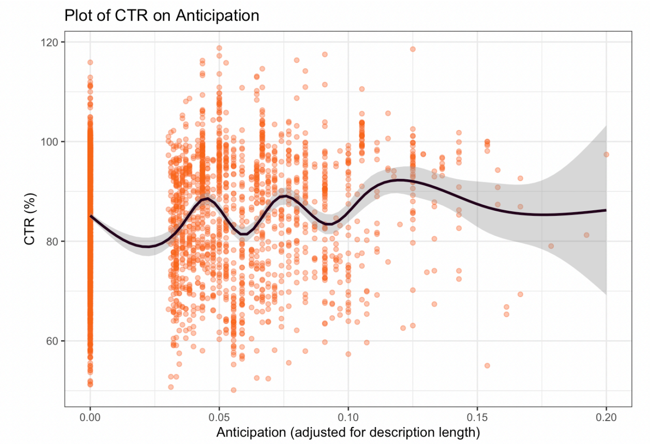
The World Bank. (2022). *GDP (current US$) - Romania*. Retrieved from World Bank Group: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=RO

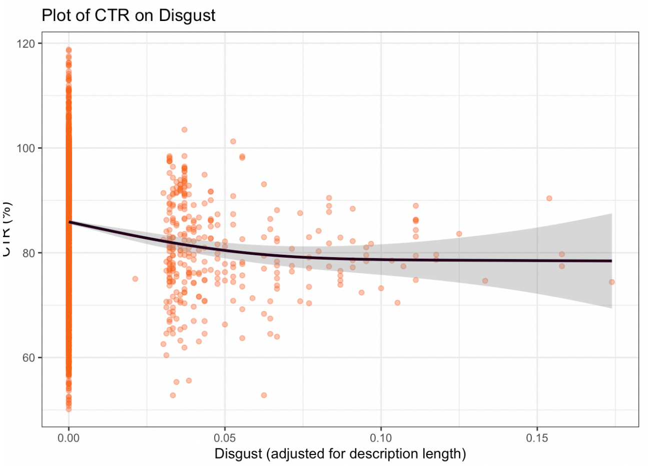
Wickham, H. (2024). *rvest: Easily Harvest (Scrape) Web Pages*. Retrieved from R package version 1.0.4, https://github.com/tidyverse/rvest: https://rvest.tidyverse.org/

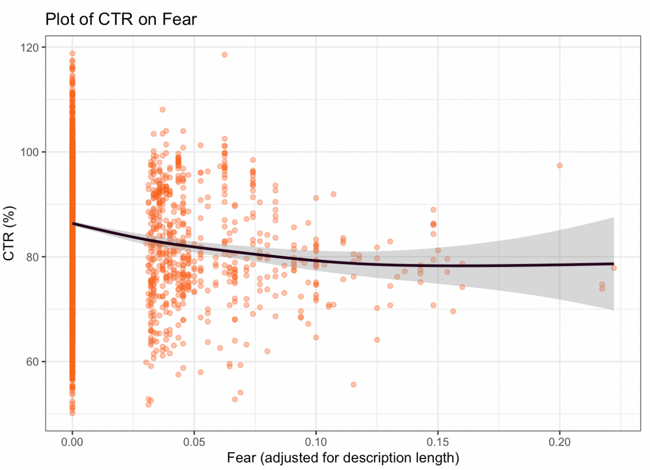
Appendix A: Further Descriptive Analysis

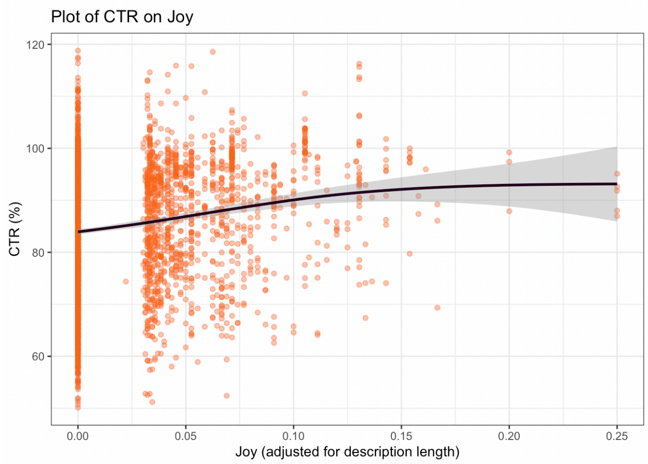


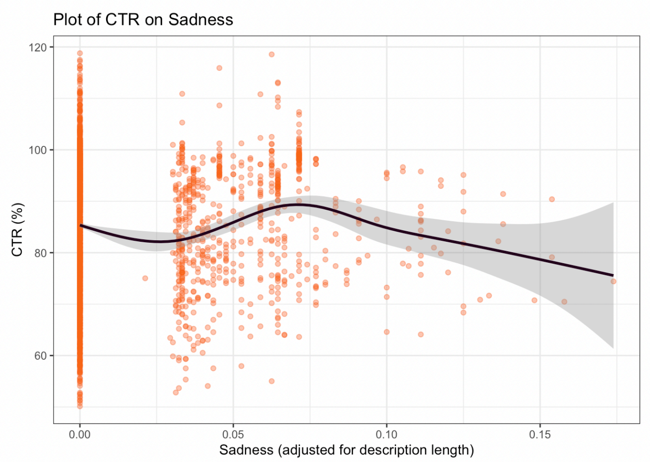


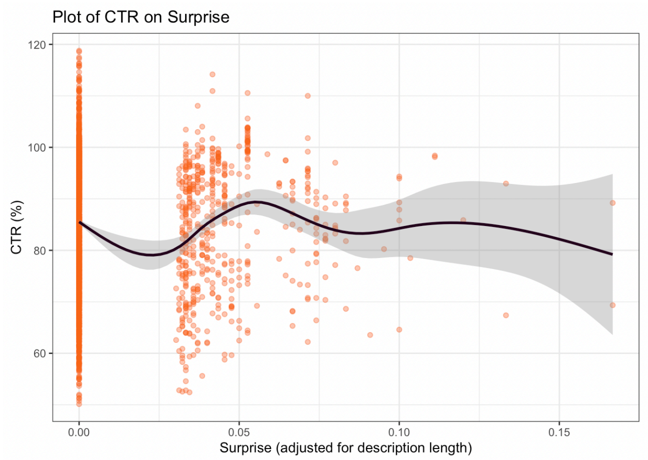


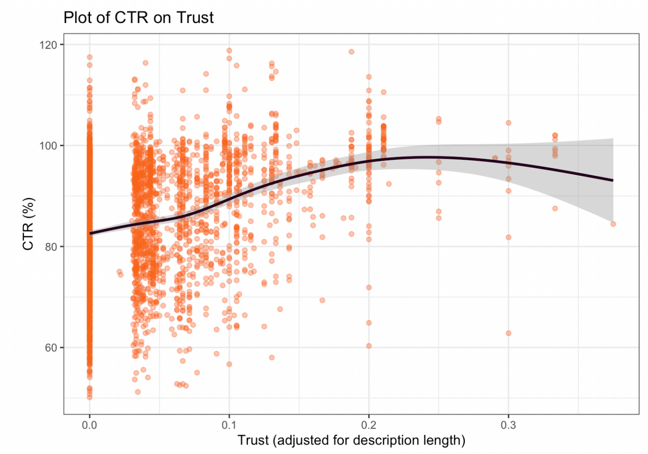


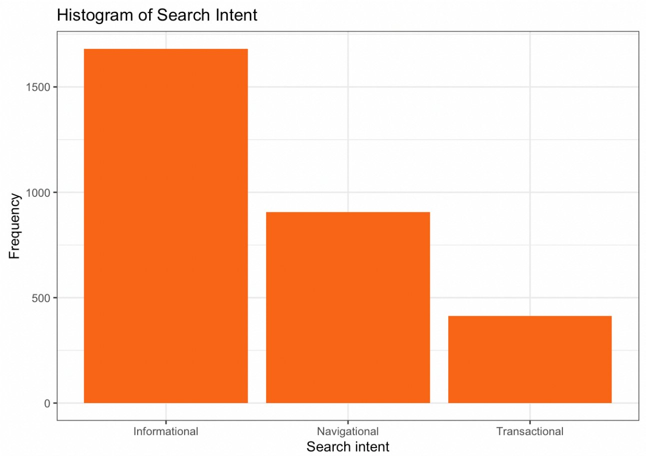


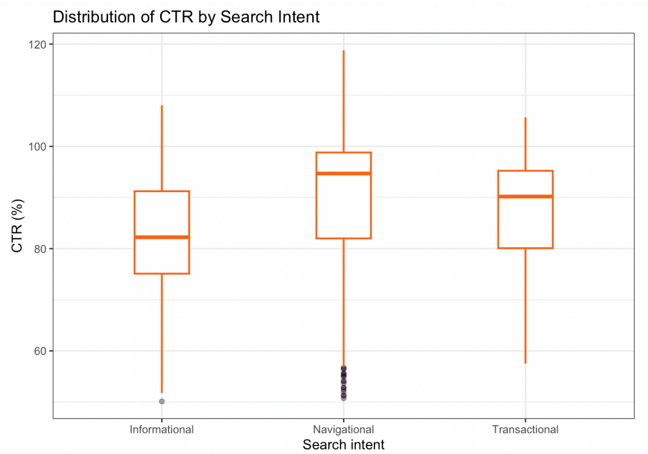


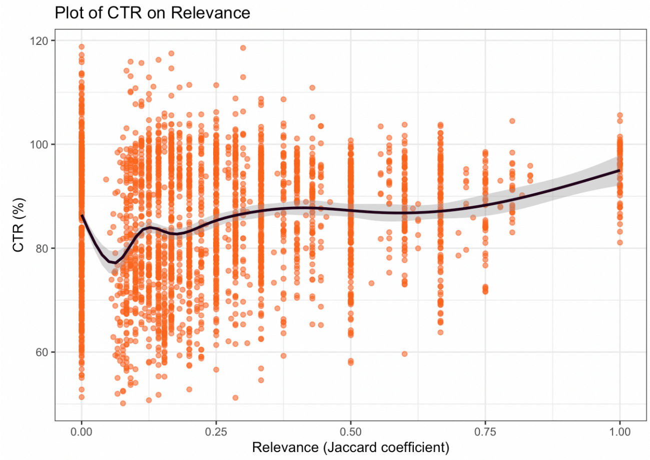




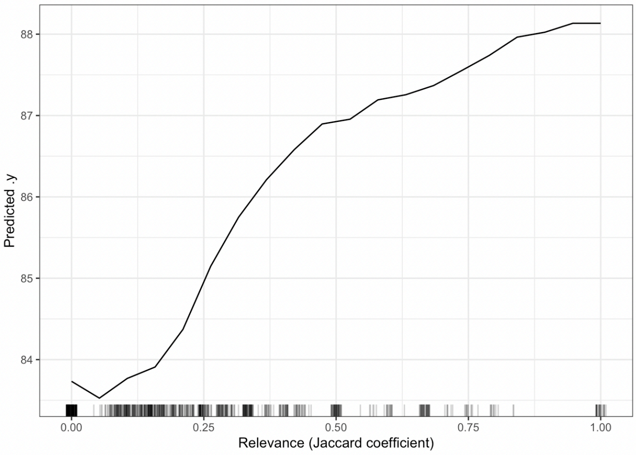

Plot of CTR on Trust
CTR (%)
Trust (adjusted for description length)

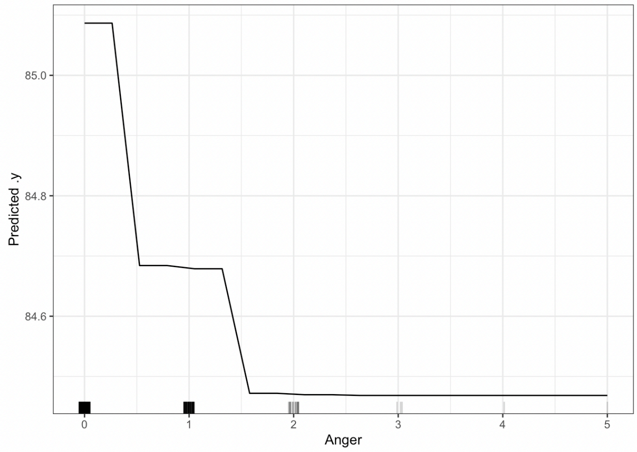

Histogram of Search Intent
Frequency
Informational
Navigational
Transactional
Search intent

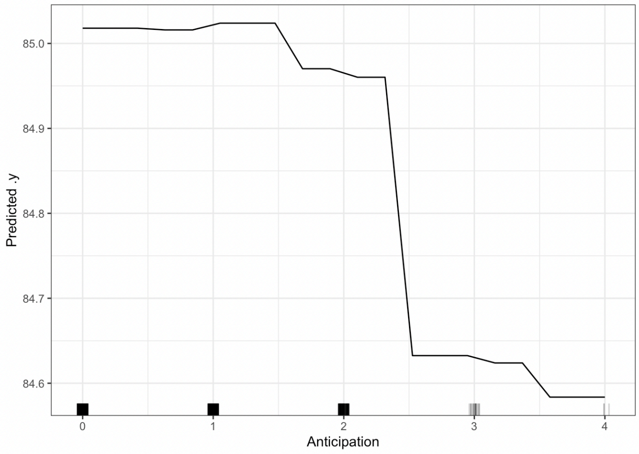

Distribution of CTR by Search Intent
CTR (%)
Informational
Navigational
Transactional
Search intent

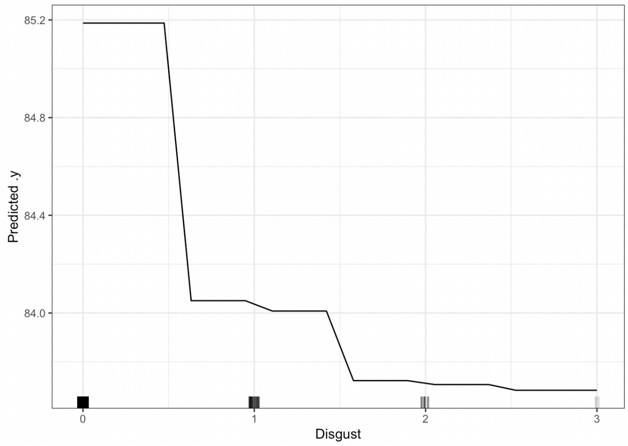

Plot of CTR on Relevance
CTR (%)
Relevance (Jaccard coefficient)

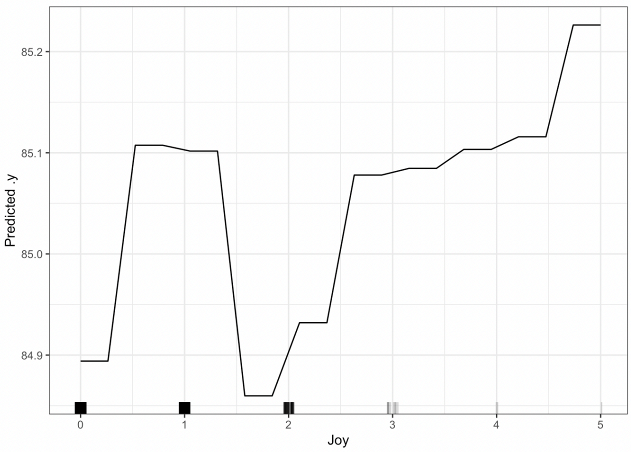
Appendix B – PDPs

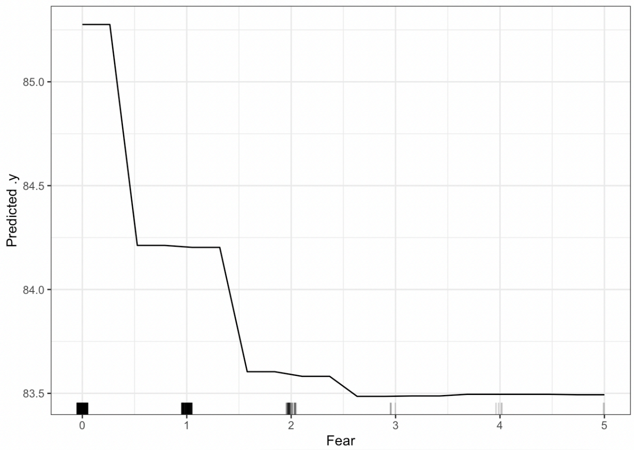


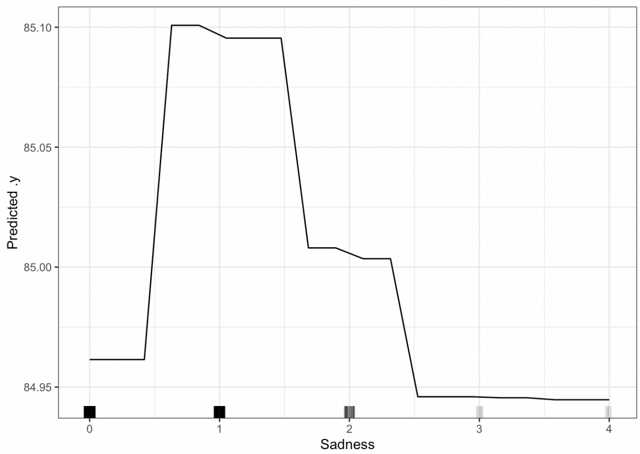


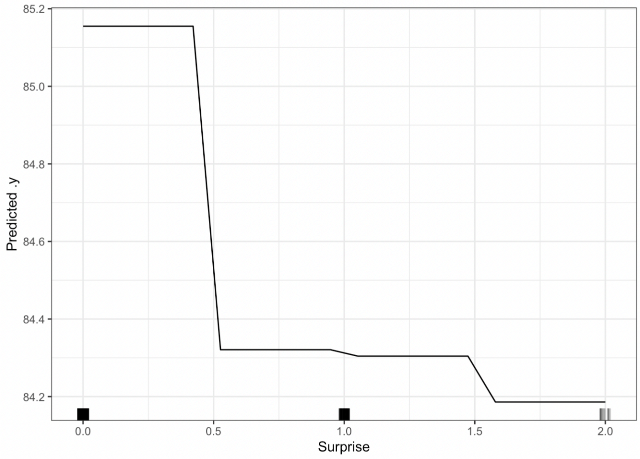


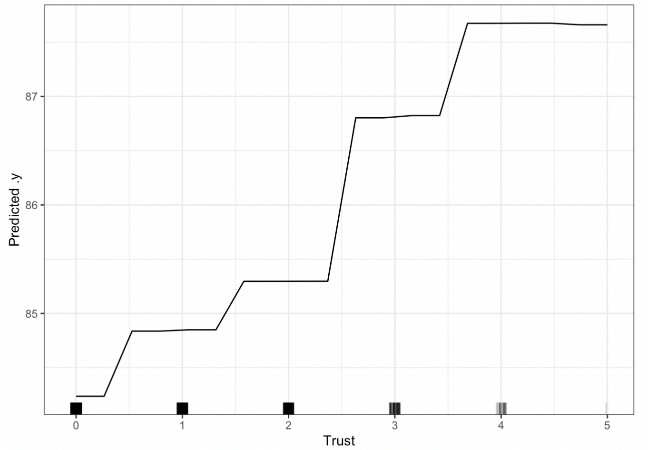


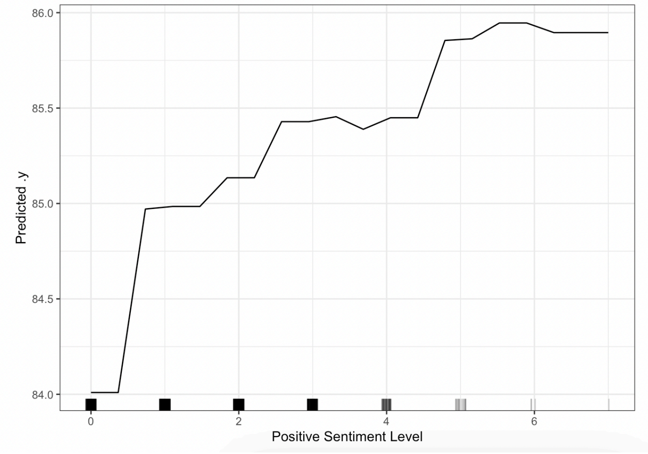


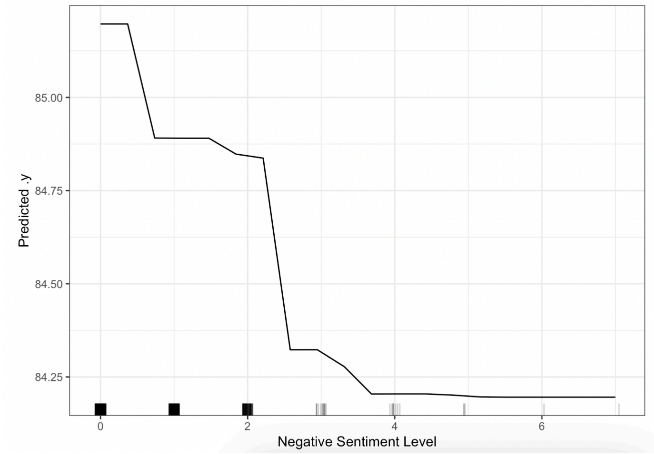






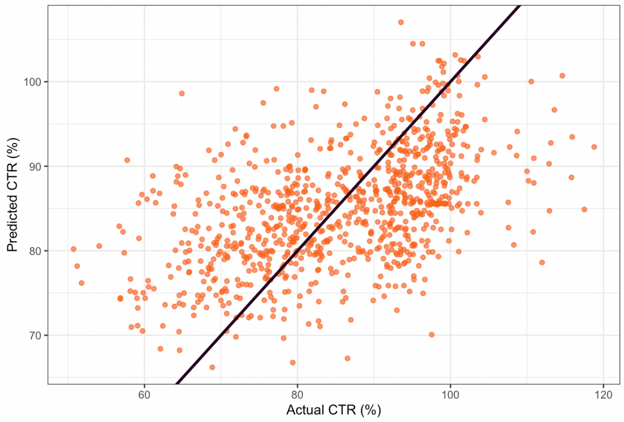

Predicted .y
85
86
87
0
1
2
3
4
5
Trust


Predicted .y
84.0
84.5
85.0
85.5
86.0
0
2
4
6
Positive Sentiment Level

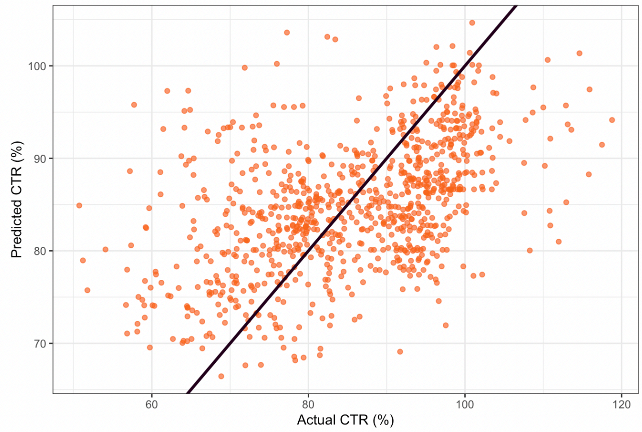

Predicted .y
84.25
84.50
84.75
85.00
0
2
4
6
Negative Sentiment Level

Appendix C – Model Performance Comparison and Diagnostics

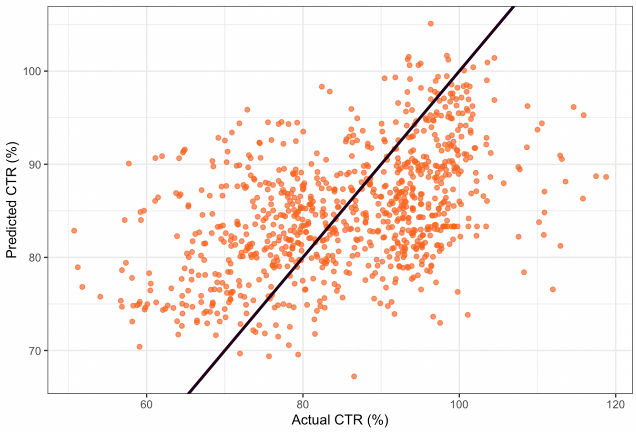
Prediction performance visualized (model: Emotion)



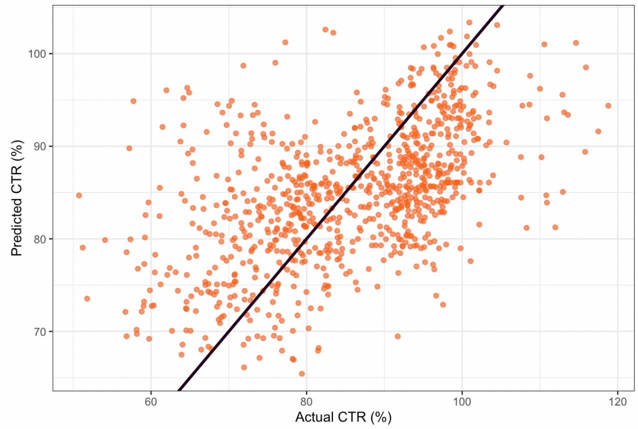
Prediction performance visualized (model: Search Intent)



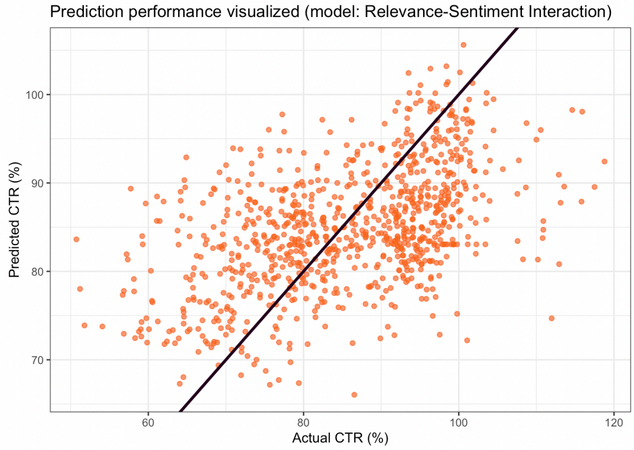
Prediction performance visualized (model: Relevance)



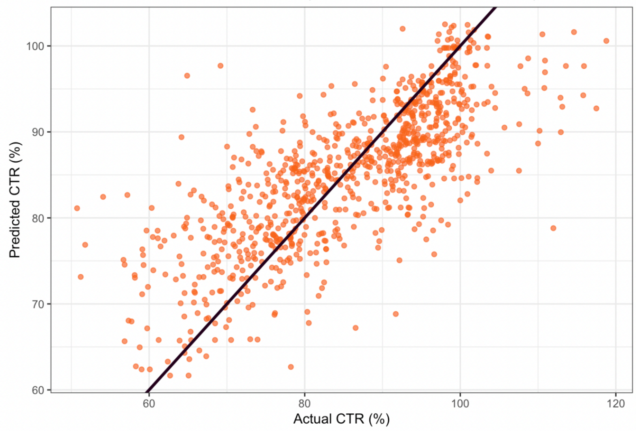
Prediction performance visualized (model: Relevance-Search Intent Interaction)



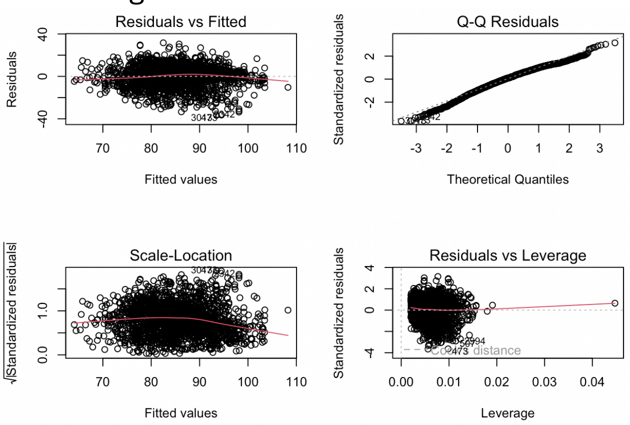
Prediction performance visualized (model: Relevance-Sentiment Interaction)



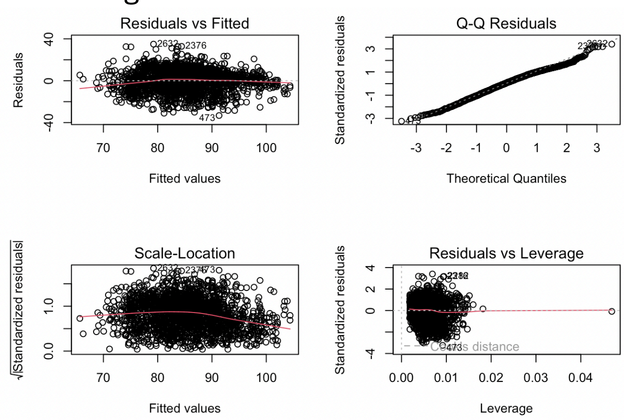
Prediction performance visualized (model: Tuned Random Forest)



Model Diagnostics – Search Intent

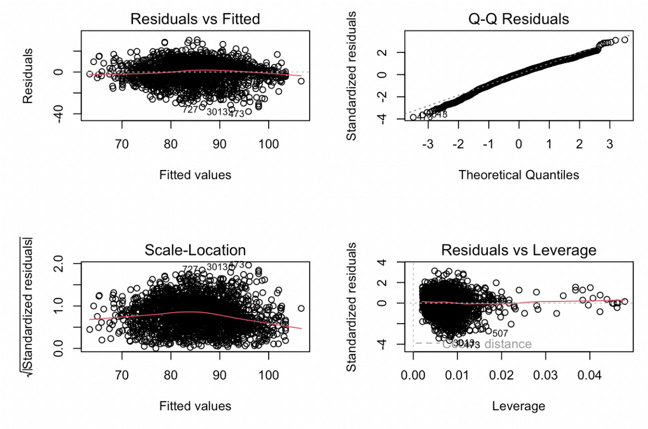


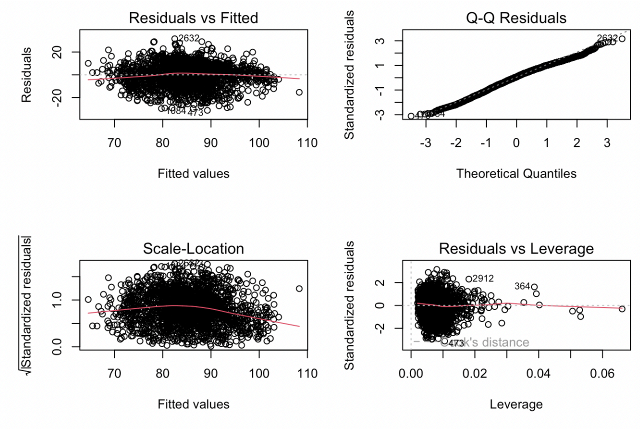
Model Diagnostics – Relevance



Model Diagnostics – Relevance:Search Intent

Model Diagnostics – Relevance:Sentiment


Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage


Residuals vs Fitted
Residuals
Fitted values
Q-Q Residuals
Standardized residuals
Theoretical Quantiles
Scale-Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized residuals
Leverage