Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia

COPPE UFRJ

A MULTI-DIMENSIONAL APPROACH TO UNDERSTANDING THE EFFECT OF PAGE CONTENT AND INFRASTRUCTURE ON PAGE LOAD TIME

Daniel Atkinson Oliveira

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientadores: Rosa Maria Meri Leão
Edmundo Albuquerque de Souza e Silva

Rio de Janeiro
Dezembro de 2024

A MULTI-DIMENSIONAL APPROACH TO UNDERSTANDING THE EFFECT OF PAGE CONTENT AND INFRASTRUCTURE ON PAGE LOAD TIME

Daniel Atkinson Oliveira

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientadores: Rosa Maria Meri Leão
Edmundo Albuquerque de Souza e Silva

Aprovada por: Prof. Rosa Maria Meri Leão
Prof. Edmundo Albuquerque de Souza e Silva
Prof. Carlos Alberto Vieira Campos
Prof. Daniel Sadoc Menasché

RIO DE JANEIRO, RJ – BRASIL
DEZEMBRO DE 2024

Atkinson Oliveira, Daniel

A Multi-Dimensional Approach to Understanding the Effect of Page Content and Infrastructure on Page Load Time/Daniel Atkinson Oliveira. – Rio de Janeiro: UFRJ/COPPE, 2024.

X, 48 p.: il.; 29, 7cm.

Orientadores: Rosa Maria Meri Leão

Edmundo Albuquerque de Souza e Silva

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2024.

Referências Bibliográficas: p. 46 – 48.

1. QoE. 2. QoS. 3. measurements. 4. PLT. I. Meri Leão, Rosa Maria *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

Acknowledgments

I would first like to thank my supervisors, professors Edmundo de Souza and Rosa Leão, for their counsel, patience, and partnership during the whole process. Their experience was essential for the development of this thesis.

I would also like to thank all of the professors and colleagues I had the privilege of meeting along the way. All of the after-class talks and lunches together made this experience one I will remember of fondly.

I would of course like to thank my parents, without whom none of this would be possible.

Lastly, I am grateful to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) and Google LLC for the financial support throughout the whole time.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

UMA ABORDAGEM MULTI-DIMENSIONAL PARA ENTENDER O EFEITO DE CONTEÚDO E INFRAESTRUTURA DE PÁGINAS NO PAGE LOAD TIME

Daniel Atkinson Oliveira

Dezembro/2024

Orientadores: Rosa Maria Meri Leão
Edmundo Albuquerque de Souza e Silva

Programa: Engenharia de Sistemas e Computação

Este estudo examina o papel crucial do *page load time* (PLT) na era digital, onde páginas web servem como fontes essenciais de informação, entretenimento e comércio. Ao reconhecer o impacto significativo do PLT na experiência do usuário, otimização de ferramentas de busca e taxas de conversão, nós abordamos não só o estado atual das métricas de complexidade de páginas, mais especificamente métricas de conteúdo e infraestrutura, via o uso de métodos estatísticos, mas também como elas afetam o PLT. Nós usamos modelos de aprendizado de máquina, tanto supervisionados quanto não supervisionados, para analisar a influência dessas métricas em múltiplos níveis: página única, categoria de página, cluster e geral. Esta abordagem multi-dimensional permite um entendimento holístico da relação entre métricas de complexidade de página e PLT. Nosso estudo conclui que a quantidade de bytes, requisições e imagens distintas são as variáveis mais importantes para predição de PLT, com o modelo por categoria de página sendo, em geral, melhor do que os outros. Também encontramos um valor de RMSE (raiz quadrada do erro médio ao quadrado) de menos de um, para a maioria dos modelos, indicando alta acurácia na estimação de PLT. Os resultados contribuem para um entendimento mais profundo dos determinantes de PLT e ajudam a encontrar caminhos para otimização de páginas web para melhor experiência de usuário e resultados de negócios. Este estudo mostra a necessidade de foco contínuo na questão de PLT no cenário dinâmico de tecnologias web e engajamento de usuários.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

A MULTI-DIMENSIONAL APPROACH TO UNDERSTANDING THE EFFECT OF PAGE CONTENT AND INFRASTRUCTURE ON PAGE LOAD TIME

Daniel Atkinson Oliveira

December/2024

Advisors: Rosa Maria Meri Leão
Edmundo Albuquerque de Souza e Silva

Department: Systems Engineering and Computer Science

This study delves into the critical role of page load time (PLT) in the digital age, where web pages serve as essential gateways to information, entertainment, and commerce. Recognizing PLT's significant impact on user experience, search engine optimization, and conversion rates, we look at not only the current state of page complexity metrics, specifically content and infrastructure, via the use of statistical methods, but also how they affect PLT. We employ both supervised and unsupervised machine learning models to analyze the influence of these metrics at multiple levels: single page, page category, cluster, and general. This multi-dimensional approach allows for a comprehensive understanding of the relationship between page complexity metrics and PLT. Our study finds that the number of bytes, requests, and distinct images are key features in PLT prediction, with the page category model generally outperforming others. We also report a root mean squared error (RMSE) of less than one in most models, indicating high accuracy in PLT estimation. The findings contribute to a deeper understanding of the determinants of PLT and offer insights into optimizing web pages for better user experiences and business outcomes. This research underscores the need for continued focus on PLT in the evolving landscape of web technology and user engagement.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

In the digital age, web pages remain an indispensable part of our everyday lives, serving as gateways to information, entertainment, and commerce. Web technology is constantly changing, as are the web development paradigms, evidenced by the debate of single-page applications *versus* multi-page applications NEOTERIC (2016) or responsive *versus* dynamically served pages GLOAGEN (2022). The use of content delivery networks (CDNs), placement of servers and the content of web pages is also evolving over time. Understanding the current state of affairs of page content and infrastructure serves as an important starting point for any analysis of page performance.

One crucial aspect influencing the user's browsing experience is page load time, commonly referred to as PLT. It is measured as the time between the beginning of the navigation and the moment the page's content begins loading, marked by the loadEventStart[1] event. This encompasses the web page's entire content, such as text, images, videos, and interactive elements. Providing a quick and smooth website is essential for guaranteeing an excellent experience for end-users, as demonstrated by substantial research from leading corporations. Over a decade ago, a study by Amazon highlighted that every 100 ms of added latency cost them 1% in sales. This discovery is not isolated; in 2006, Google noted that an additional 0.5 seconds in search page generation time led to a 20% drop in traffic GIGASPACES (July, 2023). These and other cases demonstrate why companies often invest significant resources in optimizing their websites to avoid such costly delays.

The COVID-19 pandemic has heightened the importance of delivering a high-quality user experience in web services. This global health crisis has fundamentally changed business operations, emphasizing the need for an effective online presence.

[1] https://developer.mozilla.org/en-US/docs/Web/API/PerformanceNavigationTiming/loadEventStart

Companies without an effective online presence, or those that provide low-quality digital experiences, have faced significant challenges in maintaining competitiveness and relevance in today's rapidly evolving digital world. As a site becomes less interactive, users increasingly tend to navigate away to a competitor's site. User engagement is critically linked to a website's interactivity, of which PLT is a key determinant.

PLT emerges as a pivotal factor not only in user experience but also in broader aspects such as search engine optimization and conversion rates. The significance of PLT is recognized by Google, which has integrated this metric into its search engine ranking algorithms GOOGLE (April, 2010). Moreover, there is a significant correlation between PLT and conversion rates. Pages that load within one second have been found to have conversion rates nearly 2.5 times higher than pages that take five seconds or more to load WIEGAND (April, 2022). This indicates that faster pages are more effective at converting visitors into customers. Additionally, PLT plays a crucial role in user-perceived quality. Studies like HORA *et al.* (2018) employing the Absolute Category Rating (ACR) scale, a self-reported measure of user experience, reveal a direct correlation between PLT and user satisfaction. Faster loading times are consistently associated with higher ACR scores, indicating that users perceive websites with shorter load times as more efficient and user-friendly.

In summary, enhancing PLT is not just a technical necessity but a fundamental component of delivering an acceptable user experience. It impacts various dimensions, from search engine rankings and visibility to conversion rates and overall user QoE. Therefore, understanding how page content and infrastructure influence PLT remains crucial. Previous studies exploring the relationship between page complexity metrics, namely page content and infrastructure metrics, and PLT have focused either on a few selected pages, with individual analyses being performed for each ASRESE *et al.* (2019), VOGEL and SPRINGER (2022), or on a diverse group of pages, with an overall analysis conducted for all of the pages simultaneously SAVE-RIMOUTOU *et al.* (2019), BUTKIEWICZ *et al.* (2011).

Understanding the content and infrastructure of current pages and how these influence user experiences remains crucial. This study delves into the intricate relationship between these metrics and PLT, and how it impacts user quality of experience during web browsing. Specifically, we aim to answer the following questions:

- What are the content and infrastructure trends in web pages at the present time?

- What are the most important page content and/or infrastructure features when it comes to inferring page load time?

In light of this, supervised and unsupervised models, as well as classical statistical analysis techniques are employed to answer these questions. The main contributions of this work are as follows:

- **Page characterization**: A diverse group of pages were analyzed and the findings were reported. Notably, I found the use of CDNs to be ubiquitous, with over half of the analyzed pages having all of their requests being solely to servers located in CDNs. I also found a high correlation between higher-ranked pages and number of ad services.
- **Multi-dimensional analysis**: In the multi-dimensional analysis, I examined the impact of page complexity metrics on PLT across various levels of granularity. This includes general, per-category, per-cluster, and per-page analyses, with both supervised and unsupervised models having been utilized.
- **Feature importance**: The findings reveal that the number of bytes ranks among the top three features for inferring PLT in all models under study. In the models categorized by page types, both the number of bytes and the number of requests emerged as the top two features for most categories. These observations from the page category models were consistent with findings from the individual page model, where the three most important features were identified as the number of bytes, the number of requests, and the number of distinct images.
- **Effectiveness of different models**: I individually tested the general model, the page category model, and the cluster model for each page. The page category model outperformed both the cluster and general models for ten out of the 15 pages. Notably, in most models, the root mean squared error (RMSE) was less than one, indicating a prediction error in estimating PLT of less than one standard deviation.

In the following chapters two complementary works are presented. The first, in chapter 3, reports the page characterization effort, while chapter 4 reports on the page performance work, with chapter 5 concluding.

Chapter 2

Related Work

In ASRESE *et al.* (2019) the authors conducted two longitudinal studies to analyze web page structural metrics, such as size and number of objects of a few major services. In BUTKIEWICZ *et al.* (2011) the authors selected roughly 1700 web pages and presented an overall characterization of these pages' complexity metrics, such as number of objects, size of these objects and the types of content. They also investigated the relationship between these metrics and page load times using correlation analysis as well as regression models.

In this work I performed a characterization analysis of a similar group of web pages, providing insights of the current state of affairs. For example, I found that CDNs are much more prevalent today, with the median percentage of objects fetched from CDNs being 100%. I also found that higher ranked pages have higher number of ad services.

Extensive research has thoroughly examined page performance metrics, such as page load time (PLT), time to first paint, and above-the-fold time in relation to the perceived quality of web browsing experiences. Studies involving both passive SALUTARI *et al.* (2019), HORA *et al.* (2018) and active GAO *et al.* (2017), EGGER *et al.* (2012) user interactions with pages have revealed evaluations based on various scales. These range from bad/neutral/good to a 5-point ACR scale, or involve choosing preferred pages from a group SALUTARI *et al.* (2019), HORA *et al.* (2018), GAO *et al.* (2017). Subsequently, these metrics are correlated with user scores using either expert analysis or machine learning models EGGER *et al.* (2012), HORA *et al.* (2018), HOSSFELD *et al.* (2018), JAHROMI *et al.* (2018).

In addition to studying the relationship between page performance metrics and the web users' QoE, it is also important to analyze how these metrics are influenced by various factors such as page structure/content, content provision strategies, and network conditions. AVRAM *et al.* (2014) introduces a new metric known as the latency amplification factor, which measures the impact of latency on page load times. To study this metric, they collect the page's dependency graph and estimate the overall effect on page load time through artificially added latency. In VOGEL and SPRINGER (2022), the authors report that 70% of JavaScript and 90% of CSS scripts are loaded as render-blocking code, often only utilized after the page has finished rendering. This highlights a good opportunity for optimization. For social media and news pages, our findings indicate that the number of CSS objects and the number of JavaScript objects are, respectively, the most important features for predicting page load time. This could be partly attributed to the inefficient loading of this content. In SAVERIMOUTOU *et al.* (2019), an analysis of time to first visual rendering showed lower RTTs (Round Trip Times) and fewer requests in "good response" navigations under various conditions, underscoring the significance of the number of requests in page load time prediction. Similarly, BUTKIEWICZ *et al.* (2011) finds a strong correlation between the number of bytes and PLT, while also identifying the number of requests as the best predictor of PLT.

The findings align with prior research, highlighting the importance of certain features in predicting page load time, such as the number of requests and the number of bytes. The study shows the significant impact of image-related features, particularly in the context of single pages, and the importance of the number of servers in a diverse group of pages, with respect to page load time prediction. Most importantly, it explores the relationships between complexity metrics and PLT through focused category-specific and cluster-specific analyses, providing a more comprehensive understanding of how these metrics are related than previous works, which address this problem in a much broader fashion.

Chapter 3

Page Characterization

3.1 Introduction

Understanding the fundamental characteristics of modern web pages is essential to contextualizing their performance. This chapter aims to provide an overview of the landscape of the most accessed web pages in Brazil, shedding light on how they are structured and categorized. By analyzing elements such as Multipurpose Internet Mail Extensions (MIME) content type distributions, page sizes, and per-category characteristics, we establish a foundational understanding of page complexity that serves as a basis for further exploration into page load time (PLT).

This characterization is not merely descriptive but provides valuable insights into trends and patterns that influence user experience and technical optimization. For instance, understanding how the prevalence of specific content types or the use of CDNs varies across categories can highlight factors that significantly shape PLT. This effort is a crucial step in answering broader questions about what makes web pages efficient, engaging, and performant in today's digital ecosystem.

3.2 Methodology

To understand the current state of affairs of web pages, we needed a software that would automatically browse a set of pages and collect all of the necessary metrics for that session. For this, the WebPageTest [1] framework was employed. This framework consists of a master-worker architecture, where the server schedules tasks for agents to perform. In the context of this study, the task consisted solely of navigating to a page. During the page navigation process, page-level metrics and request-level metrics were collected. The page-level metrics were collected via the navigation

[1] WebPageTest: `https://www.webpagetest.org/`

timing API [2] and the request level metrics were collected via the HAR files, which logs the browser interactions. The data of interest for this work were the request-level metrics, of which, some examples are:

- HTTP request URL
- DNS resolution time
- SSL time
- TCP connection time
- time until first byte is received
- time until last byte is received
- MIME Content type of the request
- number of incoming/outgoing bytes for the request
- HTTP request protocol (HTTP/1.1 or HTTP/2)
- flag indicating if the request url corresponds to a CDN
- HTTP request headers

The data analyzed in this chapter were gathered between December 1, 2022, and June 28, 2023. The agents used were five Raspberry Pis 4, model B, with 4GB of RAM and 1500MHz CPU with 4 cores. These Raspberry Pi devices were located at the homes of five different volunteers in three different ISPs in Rio de Janeiro, Brazil. They were equipped with custom data-collecting software [3], part of which includes the WebPageTest agent suite, and each unit was connected directly to the volunteer's gateway routers via an Ethernet cable. While it is true that Raspberry Pis are not as powerful as the average modern PC, they were used because they could be placed in a non-intrusive manner inside the volunteer's homes. While the absolute values of PLT discussed below may be affected by the use of Raspberry Pis instead of more capable machinces, all the main takeaways from this study can be generalized to all types of devices. The Ethernet connection was chosen over Wi-Fi because of the inherent variability of wireless connection conditions BISWAS *et al.* (2015).

[2]Navigation timing: `https://developer.mozilla.org/en-US/docs/Web/API/Performance_API/Navigation_timing`

[3]wptagent-automation: `https://github.com/danielatk/wptagent-automation`

For the tests, I leveraged a curated list of the top 100 most visited websites in Brazil during 2022[4], which was assembled using Clickstream data to get traffic analytics from over 200 million anonymous users. The WebPageTest server was allocated to a machine in a laboratory in UFRJ and custom software was added to automatically schedule the tasks to the agents. To ensure an equitable distribution of test executions, a sampling approach was implemented. At five-minute intervals, an agent was randomly selected from the pool of availability. Notably, the selection process followed a sampling-without-replacement strategy, meaning that each agent would be tested exactly once before being excluded from subsequent selections. This approach guaranteed that all agents had an equal opportunity to participate in the experiments before the sampling process resumed anew.

Once an agent was selected it performed a page navigation, following the same sampling-without-replacement strategy. The use or not of adblock for the test was randomly selected, as was the resolution of the screen (1366x768 or 1920x1080), which was emulated with X software, for Debian-based systems. Before the test began the agent was notified, so that it could ensure the correct adblock and resolution settings were in place. After the experiment was performed the result was sent from the WebPageTest server to a collection server.

Following the data collection phase, the focus shifted to the task of data analysis. In order to provide a comprehensive view of the state of page complexity metrics, classic statistical methods, such as correlation and distribution analysis, were employed.

3.3 Results and Discussion

For the first measurement campaign the objective was to obtain a general characterization of web pages in Brazil. This was done according to the content *vs* service complexity metrics defined in BUTKIEWICZ *et al.* (2011). Content complexity refers to the content of the web page, such as number of downloaded bytes, number of JavaScript (JS) objects and number of images. Service complexity refers to the way the content is fetched, such as number of servers and fraction of servers in CDNs. To do this the top 100 web pages in Brazil were analyzed. Figure 3.1 shows the number of web pages in each category.

Table 3.1 shows the distribution of MIME type requests and bytes. Although the percentage of video requests is the lowest of all categories, the percentage of video bytes is the third highest. In contrast, the percentage of other requests is the third highest while the percentage of bytes is the second lowest.

[4]Top 100 most accessed sites in Brazil [2022 Edition]: `https://pt.semrush.com/blog/top-100-sites-mais-visitados/`

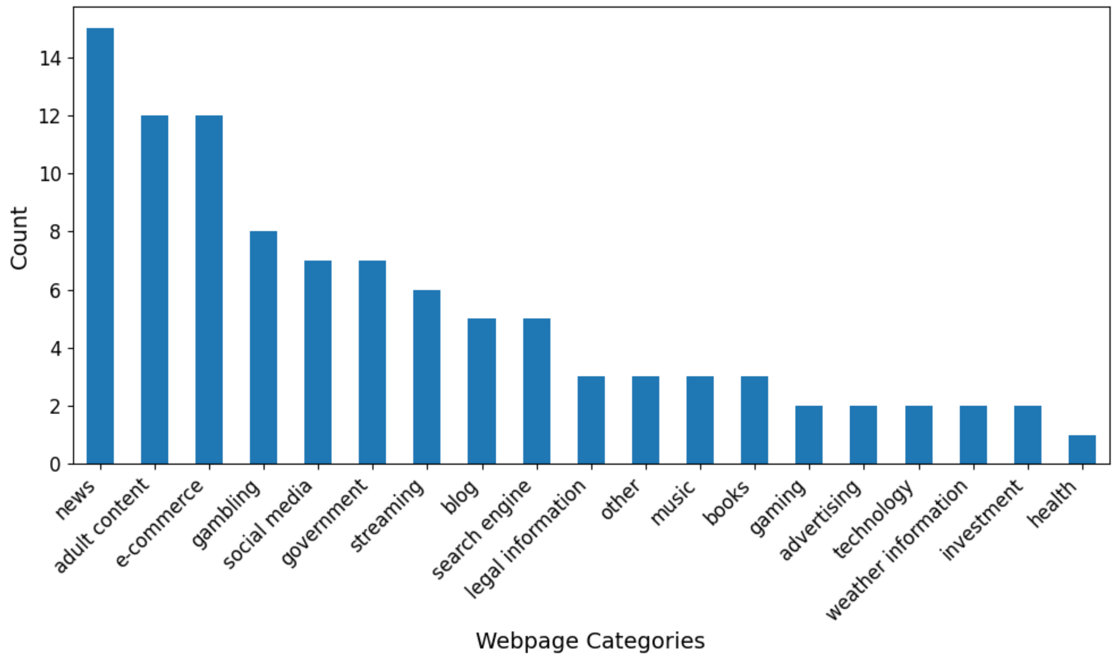

Figure 3.1: Number of pages per category

Table 3.1: Percentage of MIME types

MIME type	percentage of requests	percentage of bytes
image	41.2	47.6
video	0.3	10.6
JavaScript	23.6	30.3
CSS	3.6	1.7
html	10.0	4.0
font	2.0	3.7
other	19.4	2.1

Figures 3.2, 3.3, and 3.4 show the histograms of the number of bytes, number of requests, and number of servers, respectively. As can be seen, all three of them have heavy tails, which is true for almost all of the features. The median values of number of bytes, requests and servers are 2 million (2MB), 103, and 13, respectively.

Figures 3.5, 3.6, and 3.7 show the CDFs of the number of bytes, requests and servers, respectively, of the top five categories with most web pages. News web pages have the highest median number of servers and requests, as was found in BUTKIEWICZ *et al.* (2011). What is significantly different is how close the median values for e-commerce web pages are to news web pages in both of these metrics. This is intuitive, though, given that both are categories of web pages that usually present a lot of results in their landing pages. There is, however, a heavy tail for gambling sites, in relation to number of servers. Although the median number of bytes for both of these categories is also slightly higher than for the others, the 95th

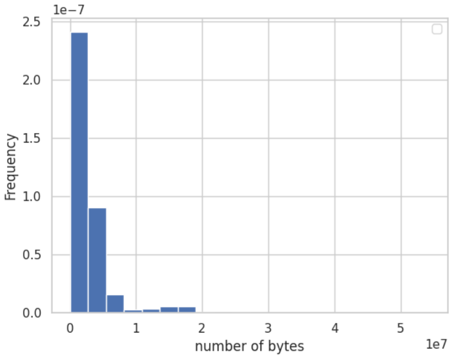

Figure 3.2: Histogram of number of bytes

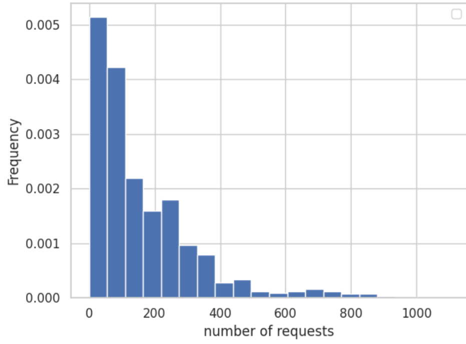

Figure 3.3: Histogram of number of requests

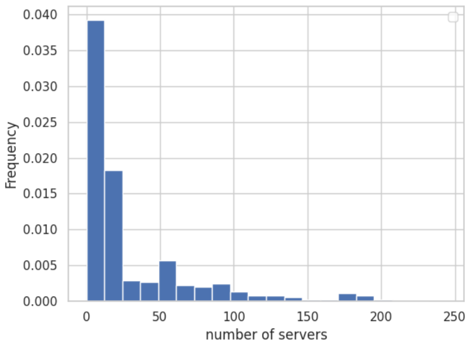

Figure 3.4: Histogram of number of servers

percentile of bytes of social media and adult content categories is much higher than that of the rest. This is due to the fact that these categories are the ones with predominantly video content, which is larger in size, as can be seen in Figure 3.8.

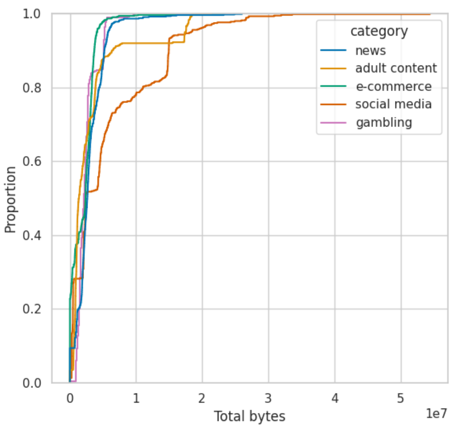

Figure 3.5: CDF of the total amount of incoming bytes for the top 5 web page categories

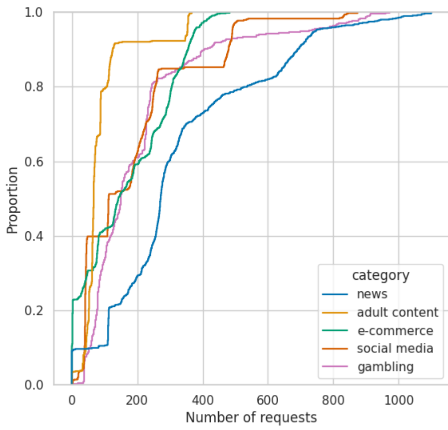

Figure 3.6: CDF of the total amount of requests for the top 5 web page categories

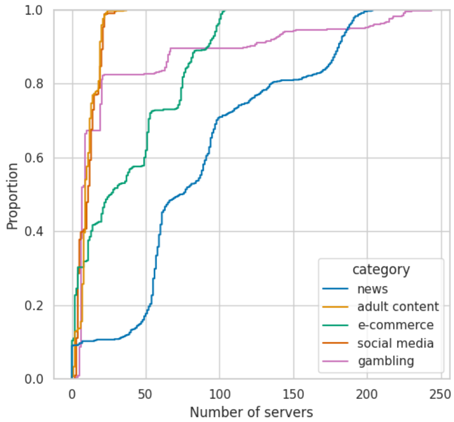

Figure 3.7: CDF of the total amount of servers for the top 5 web page categories

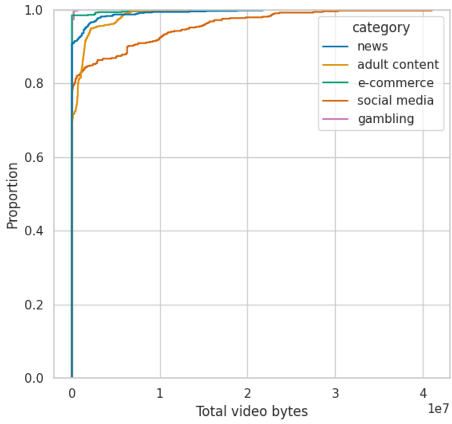

Figure 3.8: CDF of the total amount of incoming video bytes for the top 5 web page categories

Table 3.2 shows the correlation between page rank and page complexity metrics. The highest correlation found is a positive correlation between number of CDN servers and rank, with value $r = 0.213$, and in close second comes a positive correlation between number of ad services and rank, with value $r = 0.208$. One possible interpretation is that higher-ranked pages have greater advantages when it comes to the use of both CDNs and ads. They profit more with ads due to higher exposure, and CDNs should theoretically provide the page contents faster, which would result in lower page load times. It could also be the case that the ads are being served from CDNs, which would automatically correlate use of ads with use of CDNs. This is investigated further below. Although this clearly shows a trend in the top 100 pages, the variety of pages, in terms of rank position, is not very high so these results do not necessarily hold up for larger samples of rank positions.

Figures 3.9 to 3.11 show the median percentage contribution, in bytes, of the most relevant MIME types for each of the top 5 web page categories. In contrast to BUTKIEWICZ *et al.* (2011), there is no flash content, since, after Adobe discontinued support, in Chrome version 87, released December 2020, flash support was completely terminated. When compared to the 2011 study, the contribution of CSS and JavaScript, in terms of bytes, is higher. For example, in the mentioned study, the median contribution, in bytes, of JavaScript, for e-commerce and news was 39% and 34%, respectively. In this study we found contributions of 47% and 52%, respectively. For CSS, the contributions for e-commerce and news in the 2011 study were 0.07% and 0.08%, respectively. In this study, they were 1.4% and 1.0%, respectively.

Table 3.2: Pearson correlation between page rank and page complexity variables

web page variables	Correlation to rank
number of CDN servers	0.213
number of ad services	0.208
number of CSS objects	0.191
number of servers	0.185
number of font objects	0.143
number of font bytes	0.141
fraction of requests to CDN	-0.127
number of JS objects	-0.127
number of html objects	0.117
number of video objects	-0.110
number of incoming bytes from CDN	0.017
fraction of servers with CDN	-0.105
number of image objects	0.105
number of video bytes	-0.094
number of image bytes	0.091
number of CSS bytes	0.091
number of other objects	0.080
number of other bytes	-0.079
number of requests to CDN	0.054
number of requests	0.054
number of outgoing bytes	0.052
number of JS bytes	-0.046
number of html bytes	0.019
number of incoming bytes	0.017
fraction of bytes from CDN	-0.014

The contributions of image bytes showed little change. In the 2011 study the contributions for e-commerce and news categories were 36% and 27%, respectively, while in this study they were 41% and 22%, respectively. Not surprisingly, the category with the highest median percentage of image bytes is adult content.

Figures 3.12 to 3.14 show the CDFs of fraction of servers, requests, and bytes, respectively, from CDNs. While the CDF of the fraction of servers which are located at CDNs is quite diverse among the categories, the fraction of requests and bytes follows a similar pattern. In BUTKIEWICZ *et al.* (2011) the median contribution of bytes from CDNs was around 40%, however for this study the median is practically 100%, at least for the top 5 page categories. This indicates that the distribution of bytes/requests per server follows a very heavy tail, with most of the bytes/requests belonging to very few servers. This also shows that the most impactful servers are located in CDNs, for these categories. There could be a bias in the comparison, since the 2011 study used pages ranked as low as 20000 in popularity, while this one only uses the 100 most visited web pages in Brazil. Previously we observed a positive

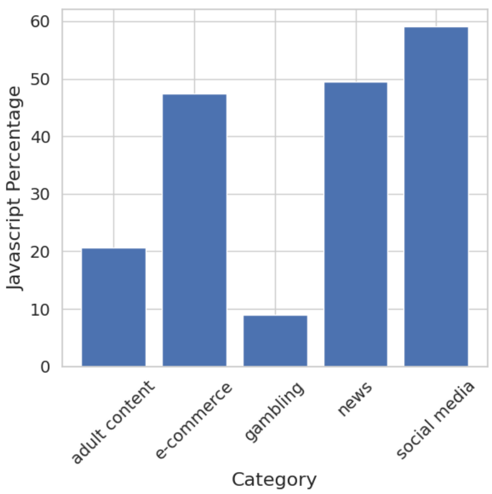

Figure 3.9: Median contribution of JavaScript bytes for the top 5 web page categories

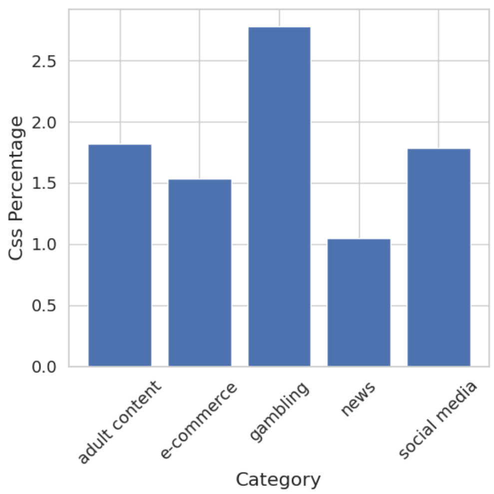

Figure 3.10: Median contribution of CSS bytes for the top 5 web page categories

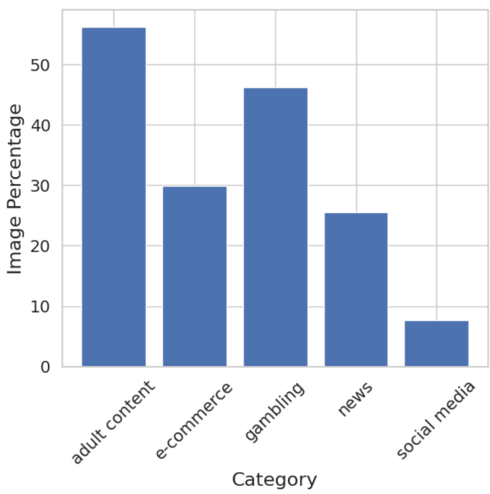

Figure 3.11: Median contribution of image bytes for the top 5 web page categories

correlation between use of CDN and rank. It is possible that the lower-ranked pages analyzed in the 2011 study reduced the overall percentage use of CDNs. Whether or not that is the case, the large difference in results indicates that the use of CDNs has become more prevalent in recent years.

Figure 3.15 shows the CDF of the total amount of distinct ad services for the top five web page categories. As can be seen news web pages have the highest value for almost all the percentiles. This corresponds to the finding of the news category having a high percentage of JavaScript bytes and requests. While for e-commerce and social media, which also have high percentages of JavaScript bytes, these scripts are probably used for deliberately designed dynamic elements of the page; for news web pages these scripts are probably mostly used to load ad content.

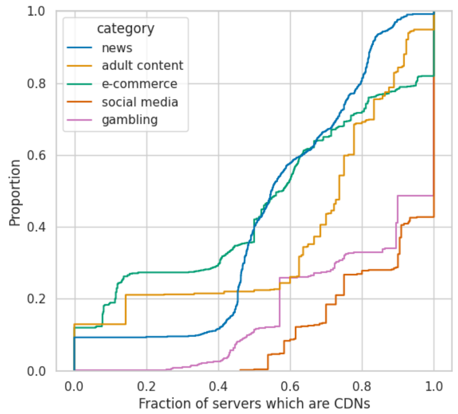

Figure 3.12: CDF of the fraction of servers which correspond to CDNs for the top 5 web page categories

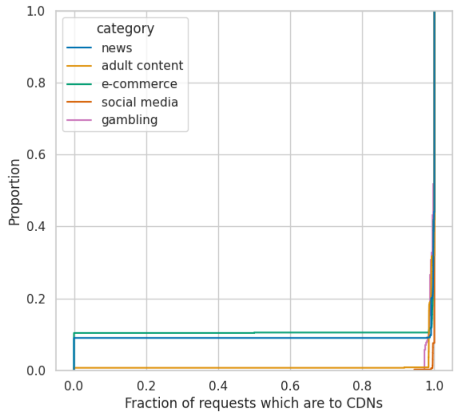

Figure 3.13: CDF of the fraction of requests which are to CDNs for the top 5 web page categories

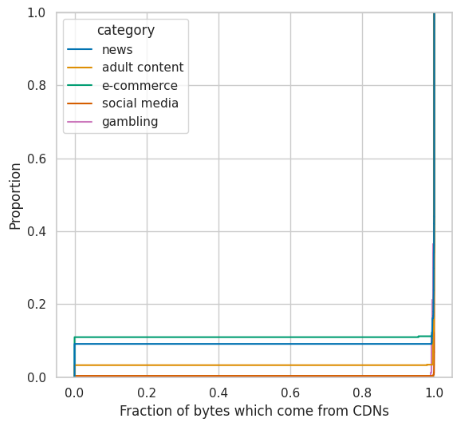

Figure 3.14: CDF of the fraction of bytes which come from CDNs for the top 5 web page categories

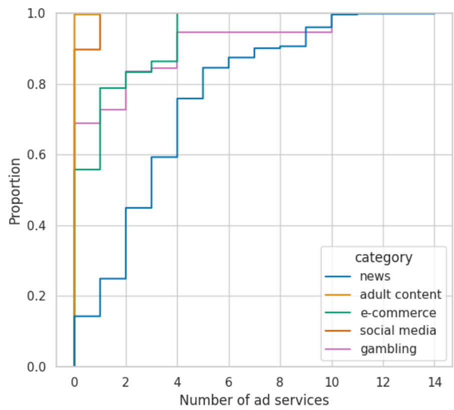

Figure 3.15: CDF of the total amount of ad services for the top 5 web page categories

Adblock and screen resolution were toggled in order to analyze the effect of these differences. Point-biserial correlation was calculated for each of these factors in relation to the page complexity metrics but no significant correlation was found. The same analysis was performed with the top five page categories, also did not result in any significant correlations.

To understand the correlation between the complexity variables a correlation matrix was generated, shown in Figure 3.16. The high correlation (0.77) between incoming bytes and image bytes reinforces the fact that most bytes are from images. Similarly, there is a high correlation between number of requests and number of image objects (0.80). In contrast to this, the correlation between incoming bytes and number of servers is 0.06, which shows the high variability in number of bytes between different numbers of servers. Number of image objects has a positive correlation with number of objects of all other types, except for video, with which it seems not to have any correlation. The reason for this, however, is not clear. As expected, number of JavaScript bytes and objects both have a positive correlation to the number of ad services, of 0.52 and 0.36, respectively.

Interestingly, there is a correlation of 1 between number of requests to CDN and number of requests as well as between number of incoming bytes from CDN and number of incoming bytes. This is in line with the previous findings of the median percentage of requests/bytes being to/from servers located in CDNs. The highest correlation to number of ad services is to number of CDN servers, indicating that most ad content is hosted in CDN servers.

The correlation between number of image bytes and image objects, of 0.39, is relatively small when compared to the same correlation for other types of content. This indicates a high variability in the size of images. Furthermore, number of ad services has a negative correlation with number of image bytes while having a high positive correlation (0.57) to number of image objects, indicating that much of the ad content is images, but of a small size.

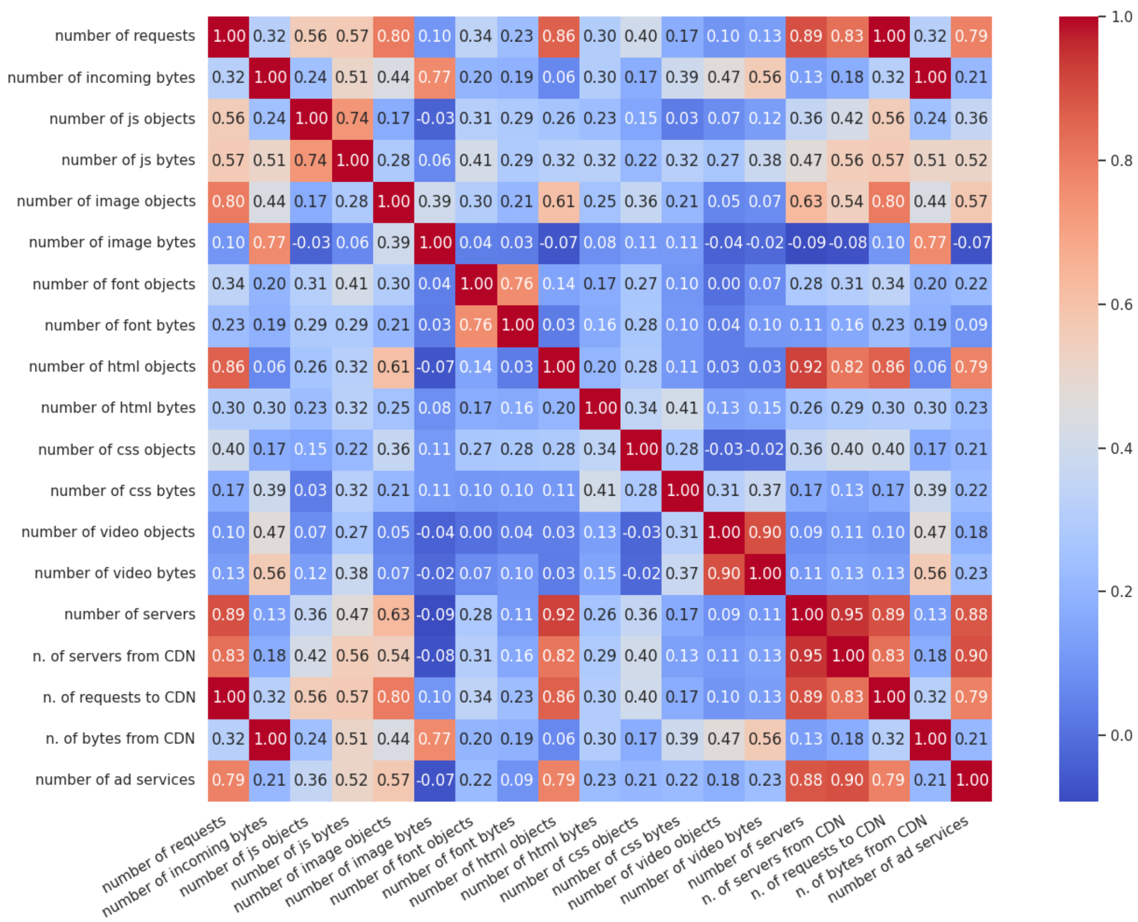

Figure 3.16: Correlation matrix of complexity metrics

Chapter 4

Page Performance

4.1 Introduction

The efficiency with which a web page loads is critical in shaping user experience and business outcomes, making PLT a focal point of web optimization efforts. Building upon the insights from the previous chapter, this chapter delves into the intricate relationship between page complexity metrics—such as the number of requests, number of bytes, and number of servers—and PLT.

The analysis presented here is multifaceted, examining PLT from several perspectives, including individual pages, categories, and clusters, to uncover both universal and context-specific determinants of performance. By identifying the key metrics that influence PLT and testing their predictive power across various models, this chapter provides actionable insights into optimizing web performance. The findings are not only theoretical but also practical, offering a roadmap for web developers and businesses to prioritize features that drive faster load times and better user satisfaction.

4.2 Methodology

To understand the relationship between page complexity metrics and PLT, I needed a software that could scale in a way that would allow multiple measurement nodes to test in parallel. For this, instead of WebPageTest, Node Puppeteer was used, instrumenting the Google Chrome browser, which was loaded with a plugin developed in HORA *et al.* (2018). This plugin collects important page loading timing information as well as page complexity metrics and was modified to send these data to our collection server. The page complexity metrics collected include:

- number of servers;
- number of bytes;

- number of requests;
- number of CSS objects;
- number of JS objects;
- number of image objects;
- number of distinct images;
- number of image pixels.

The data that was analyzed in this study were gathered between July 5 and September 18, 2023. The test nodes used were the same five Raspberry Pis of the first measurement campaign, instrumented with custom data-collecting software [1].

The structure of some pages can change frequently, so, to get a statistically significant number of samples for each page structure profile, the testing interval was sampled from an exponential distribution with a mean of 30 minutes. For each test, all the selected pages were navigated to, sequentially. This limited the number of pages that could be analyzed, which was defined as 20. If a page was not fully loaded within 18 seconds, the navigation was cancelled and no data were gathered for that page in that test instance. This methodology was adopted in SAVERIMOUTOU *et al.* (2019). This means that each test would have a maximum duration of around 400 seconds, just under seven minutes. The order the pages were selected followed a sampling-without-replacement strategy, meaning that each page would be tested exactly once before being excluded from subsequent selections.

The 20 pages were selected from a curated list of the top 100 most visited websites in Brazil during 2022[2]. These pages were chosen based on the following criteria:

- preference for higher ranked pages;
- broad representation of web page categories;
- preference for pages where the landing page contains the content users usually consume on the site.

One limitation of this study is that it only uses landing pages and not internal web pages. This issue was addressed by AQEEL *et al.* (2020), who found that two thirds of the analyzed papers required at least minor revisions due to this limitation. The final selection criterion was chosen to partially mitigate this issue, with the expectation that web pages with similar landing and internal pages would be less

[1] wptagent-automation: `https://github.com/danielatk/wptagent-automation`

[2] Top 100 most accessed sites in Brazil [2022 Edition]: `https://pt.semrush.com/blog/top-100-sites-mais-visitados/`

impacted by this limitation. Out of the 20 selected pages, 18 received sufficient navigation data on each of the Raspberry Pis. The resulting 18 pages, along with their respective ranks and categories, are presented in Table 4.1. From the top 34 most visited pages, we selected 18 based on the criteria described above. In this group of 18 pages, the representation of each category was significant, with categories being represented at levels ranging from 50% to 100% of those found in the list of the top 34 pages.

Table 4.1: Pages Tested

Rank	Page	Category
21	123movies.net	streaming
19	amazon.com.br	e-commerce
30	americanas.com.br	e-commerce
11	caixa.gov.br	government
4	globo.com	news
35	letras.mus.br	music
26	magazineluiza.com.br	e-commerce
14	mercadolivre.com.br	e-commerce
28	olx.com.br	e-commerce
7	pornhub.com	adult
22	reddit.com	social media
29	shopee.com.br	e-commerce
18	spankbang.com	adult
32	tiktok.com	social media
34	twitch.tv	streaming
9	twitter.com	social media
5	uol.com.br	news
3	xvideos.com	adult

Following data collection, the next phase was data analysis. To determine the most critical page complexity metrics for inferring PLT, various feature importance methods were utilized. Specifically, these methods included recursive feature elimination, forward and backward sequential feature selection, Gini importance, and permutation importance. Each of these methods was applied in conjunction with traditional regression models. Interpretable regression models, such as decision trees and random forests, were also employed, offering additional insights into the significance of the selected features. The analysis was conducted at different levels of resolution: individual page level, page category level, and general level. Figure 4.1 shows a representations of these different models.

Additionally, an unsupervised approach was employed to cluster the pages into groups based on similar page complexity metrics, with analyses also performed at this level of resolution. This approach involved using tensor decomposition to identify the principal relationships among the page complexity metrics in a low-dimensional space. Following this, the k-means algorithm was applied to cluster the

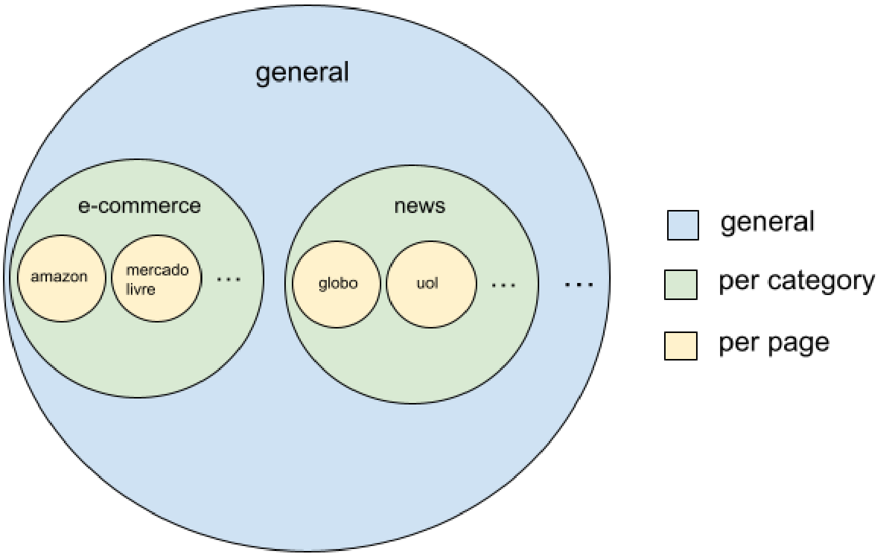

Figure 4.1: Granularity of different models

data, enabling the impact of complexity metrics on PLT within each cluster to be examined.

4.3 Results and Discussion

4.3.1 Feature Importance

As a first step to understanding the effect of complexity metrics on PLT, feature importance was analyzed. This gave an understanding of which metrics actually had a significant impact on PLT. For this, the following methods were used: Pearson correlation, recursive feature elimination (RFE), forward and backward sequential feature selection (FSFS and BSFS, respectively), Gini importance, and permutation importance.

Linear regression, decision tree, and random forest models were employed, complemented by the various feature selection methods. For each combination of model and feature selection method, 5-fold validation was used. In this process, the data were divided into training and testing groups, with 80% used for training and 20% for testing in each fold.

Table 4.2 displays the features selected by each method. The Gini importance method, which was employed together with the random forest model, selected three features, by calculating the RMSE of the estimated PLT for each number of features and applying the elbow method, as is shown in Figure 4.2. The permutation importance method, also employed together with the random forest, selected four features, via the same criteria as Gini importance, as can be seen in Figure 4.3. The number of features chosen for the Pearson correlation coefficient method was three, independent of the model being evaluated. The RFE, FSFS, and BSFS methods utilized all features, except in the case of the decision tree combined with the BSFS method, where only four features were selected. Number of servers was the only feature selected by all methods. Interestingly, while number of JS objects and number of image pixels have two of the three highest Pearson correlation values, they are not as important as other features when non-linear relations are investigated, implying there are strong non-linear relationships that can be leveraged by the more complex models.

Table 4.2: Feature importance per method

feature	Gini	Perm.	Corr.	RFE	FSFS	BSFS+LR	BSFS+DT
bytes	**0.149**	**1.143**	-0.167	**1**	**1**	**1**	0
distinct images	0.087	**0.965**	0.232	**1**	**1**	**1**	0
servers	**0.321**	**0.800**	**0.488**	**1**	**1**	**1**	**1**
requests	0.066	**0.462**	0.125	**1**	**1**	**1**	**1**
CSS objects	**0.166**	0.354	0.102	**1**	**1**	**1**	**1**
JS objects	0.100	0.298	**0.501**	**1**	**1**	**1**	**1**
image pixels	0.094	0.388	**-0.283**	**1**	**1**	**1**	0
image objects	0.016	0.214	-0.097	**1**	**1**	**1**	0

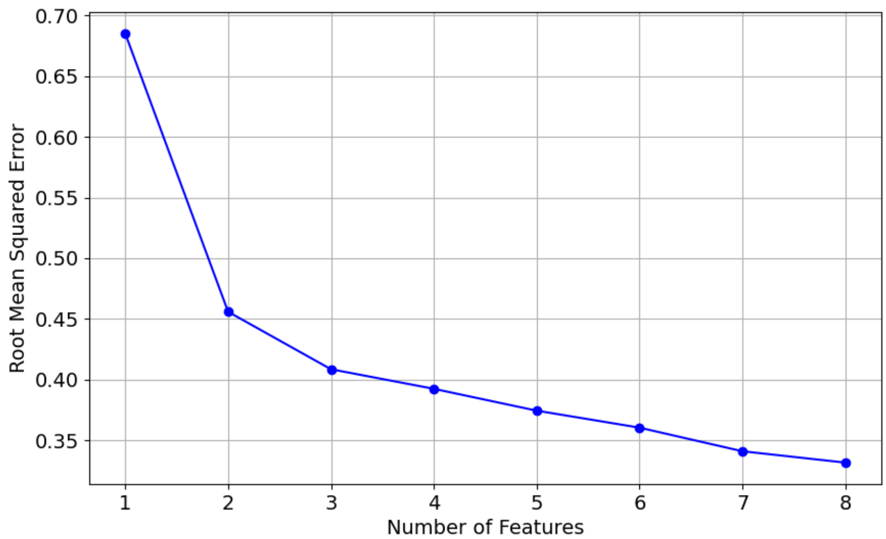

Figure 4.2: RMSE of estimated PLT by number of features selected via Gini importance for random forest model

The RMSE and mean absolute error (MAE) of the PLT predictions for all the models are shown in Table 4.3. For linear regression the feature selection performed with recursive feature elimination yielded the best results. For decision tree regression the results using the different feature sets were all very similar, with the lowest

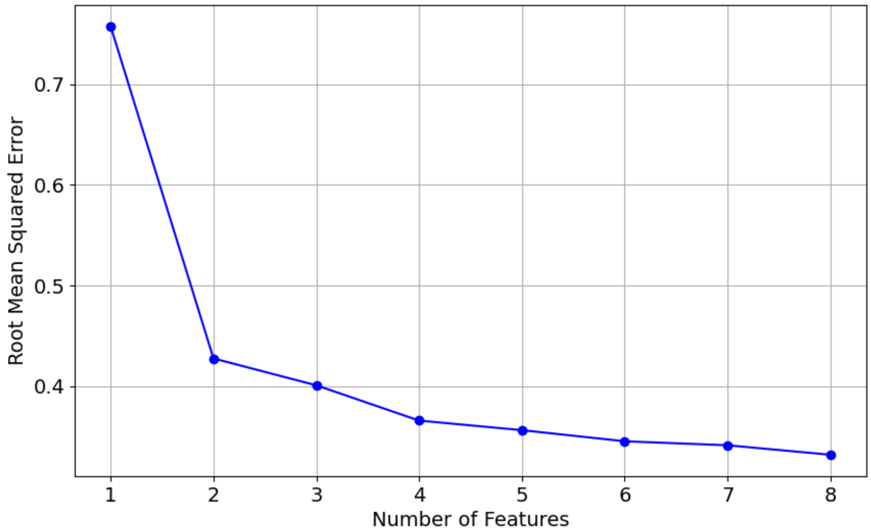

Figure 4.3: RMSE of estimated PLT by number of features selected via permutation importance for random forest model

Table 4.3: Results of PLT prediction using different models

Model	Feature Set	RMSE	MAE
Linear Regression	Gini	0.814	0.645
Linear Regression	Perm.	0.854	0.673
Linear Regression	Corr.	1.001	0.909
Linear Regression	RFE	0.739	0.575
Linear Regression	FSFS	0.739	0.575
Linear Regression	BSFS	0.739	0.575
Decision Tree Regression	Gini	0.483	0.260
Decision Tree Regression	Perm.	0.424	0.219
Decision Tree Regression	Corr.	0.410	0.235
Decision Tree Regression	RFE	0.423	0.218
Decision Tree Regression	FSFS	0.423	0.218
Decision Tree Regression	BSFS	0.412	0.214
Random Forest	Gini	0.409	0.230
Random Forest	Perm.	**0.366**	**0.204**
Random Forest	Corr.	0.415	0.232

MAE obtained when using BSFS and the lowest RMSE obtained when using the correlation features. With random forest the results were comparable whether using correlation metrics or features identified as most important by the other methods. The best-performing model out of all of them, in relation to RMSE and MAE, was random forest, using the feature set obtained via permutation importance. The features selected via this method were number of bytes, number of distinct images, number of servers, and number of requests.

The interpretation of random forest models is quite limited given that it is an ensemble model that uses many individual decision trees. One way of circumventing this is through the use of a global surrogate; i.e., a simpler, usually more interpretable model that can be used to reproduce the behaviour of a more complex model MOLNAR (2019). In this case, a decision tree model was employed as a global surrogate for the random forest model. This was done by training the decision tree with the original training set, but replacing the labels with the output of the random forest model. The tree for the surrogate model, using the features selected via Gini importance, is shown in Figure 4.4. The maximum depth was set at 2 for better visualization. The root divides the tree according to the number of servers, using 16 as the cut-off point, with two thirds of the samples being below that. For those samples, the next discriminator is number of CSS objects, with 47 as the cut-off point, with just under a tenth of the samples being above that. The samples with higher number of CSS objects had an average PLT roughly double that of the samples with lower number of CSS objects. Counterintuitively, samples with number of servers above 16 and number of kBytes above 1086 resulted in a low PLT of around 3.6 seconds. However, this occurred in only a small fraction of the samples (approximately 0.2%), indicating that such instances should be considered special cases.

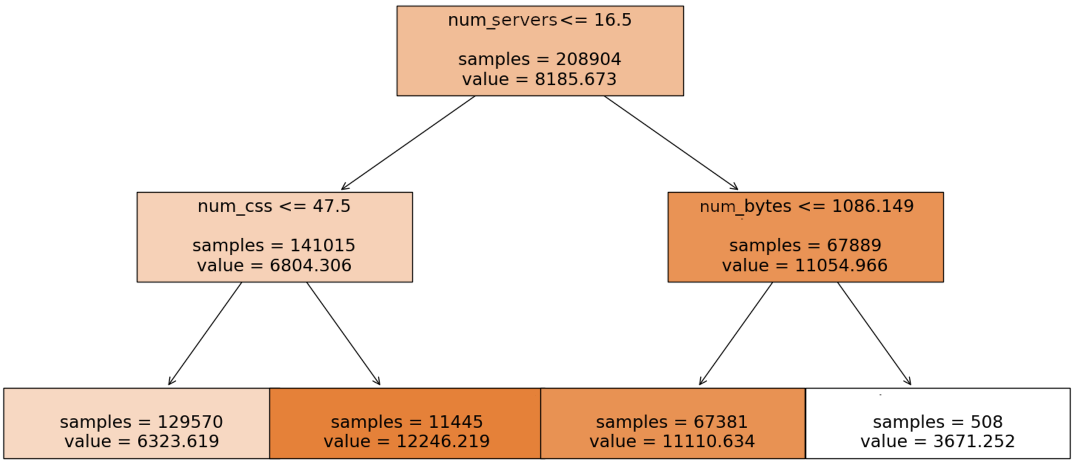

Figure 4.4: Surrogate decision tree for random forest model

Analogously, the tree for the surrogate model, using the features selected via permutation importance, is shown in Figure 4.5. As was the case with the other surrogate decision tree, number of servers was the root node, with the same value split. However, for this tree, number of requests and number of distinct images were on the next tree level, with the cut-off points being 75 and 157, respectively. For both cases the split does not create a great distinction between PLT values, with more fine-grained differences probably appearing at deeper levels.

The same models were applied in a per-page fashion and, as was the case in the general analysis, for most pages random forest was the best-performing model.

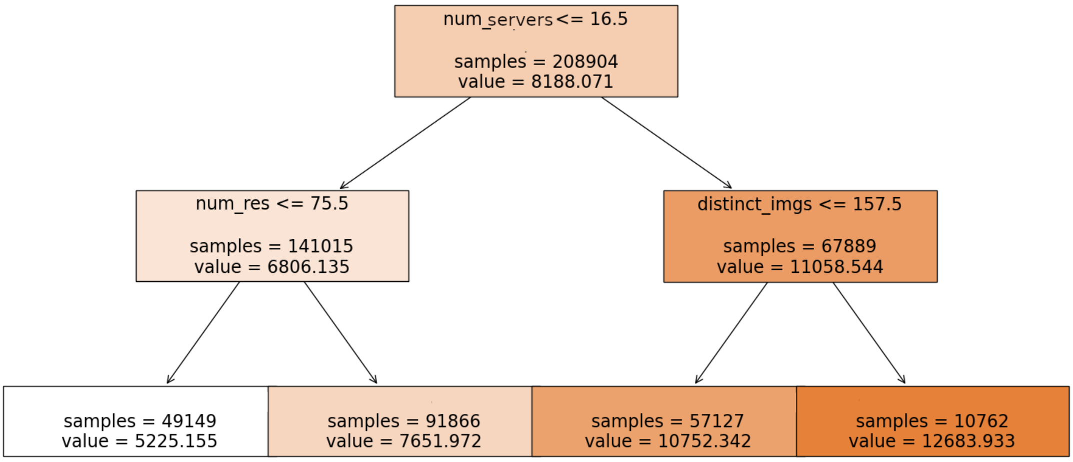

Figure 4.5: Surrogate decision tree for random forest model with permutation importance

Table 4.4 presents the top three features selected by the random forest model for each page, along with their corresponding Gini importance values, while Table 4.5 presents the same analysis, but with permutation importance.

As observed, number of bytes and number of requests were the features chosen in most cases, with these being prominent in 14 out of the 18 pages, when analyzing according to Gini importance and in 11 out of the 18 pages when analyzing according to permutation importance. In 11 out of these 18 pages, when analyzing according to Gini importance, the features including number of bytes, number of requests, and one image-related aspect (such as number of image objects, number of distinct images, or number of image pixels) were selected. The same effect was found in 7 out of the 18 pages, when analyzing according to permutation importance. In both analyses, only five pages did not include an image-related feature. Also, number of distinct images was chosen more frequently than number of image pixels and number of image objects, suggesting that it may better represent the impact on PLT compared to other image-related features.

Table 4.6 shows the comparison, in relation to RMSE, between Gini importance and permutation importance feature selection methods, in the per-page analysis. Out of the 18 pages, Gini importance was better than permutation importance in 8 pages, permutation importance was better than Gini importance in 7 pages and both performed equally well in 3 pages, with neither of the methods clearly outperforming the other. The worst performance was for the Reddit page, with an RMSE value of 0.541, which is around half a standard deviation. It must be noted that, for this analysis, the normalization was applied at a per-page basis, so results are relative to the ranges of the particular page being scrutinized.

The analysis was also conducted on a per-category basis, yielding the results, according to Gini importance, presented in Table 4.7, and according to permutation

Table 4.4: Selected Features by Random Forest per Page with Gini Importance

web page	kBytes	servers	imgs.	JS	CSS	requests	dist. imgs.	pixels
reddit	0.110	—	—	—	—	0.187	**0.513**	—
magazineluiza	**0.395**	—	0.236	—	—	0.121	—	—
pornhub	**0.707**	—	—	—	—	0.049	0.166	—
globo	0.184	—	—	—	—	**0.490**	—	0.095
shopee	0.139	—	0.072	—	—	**0.659**	—	—
olx	**0.661**	0.176	—	—	—	—	0.039	—
twitch	0.361	—	—	0.090	—	**0.364**	—	—
xvideos	**0.345**	—	—	0.168	—	0.151	—	—
letras	0.184	—	—	**0.458**	—	—	0.168	—
americanas	0.155	—	—	—	—	0.208	—	**0.287**
caixa	**0.773**	—	—	—	—	0.098	—	0.037
tiktok	—	—	—	**0.251**	0.194	0.141	—	—
amazon	—	—	0.170	—	—	0.147	**0.343**	—
mercadolivre	**0.394**	—	—	—	—	0.172	0.156	—
123movies	0.245	—	—	—	—	**0.468**	0.110	—
uol	**0.358**	—	—	—	—	0.126	—	0.180
spankbang	**0.619**	0.092	—	—	—	0.156	—	—
twitter	**0.580**	—	—	0.059	—	0.267	—	—

Table 4.5: Selected Features by Random Forest per Page with Permutation Importance

web page	kBytes	servers	imgs.	JS	CSS	requests	dist. imgs.	pixels
reddit	0.322	—	—	—	—	0.334	**0.608**	—
magazineluiza	0.650	—	**1.016**	—	—	0.576	—	—
pornhub	**0.967**	—	—	—	—	0.469	0.898	—
globo	0.370	—	—	0.621	—	**1.974**	—	—
shopee	—	—	**0.924**	—	0.276	0.794	—	—
olx	**1.685**	0.310	—	—	—	—	—	0.203
twitch	0.530	—	—	—	0.662	**1.920**	—	—
xvideos	**0.938**	—	—	0.804	—	—	0.195	—
letras	0.256	—	0.316	**0.875**	—	—	—	—
americanas	0.282	—	—	—	—	0.438	—	**1.001**
caixa	**1.026**	—	—	—	—	0.488	0.333	—
tiktok	—	—	—	**0.726**	0.615	0.606	—	—
amazon	—	—	**0.814**	—	—	—	0.446	0.453
mercadolivre	**0.851**	—	0.588	—	—	—	—	0.726
123movies	0.323	—	—	—	—	**1.593**	0.499	—
uol	**0.789**	—	—	—	—	0.438	—	0.390
spankbang	0.889	0.571	—	—	—	**0.900**	—	—
twitter	**1.551**	0.279	—	—	—	1.087	—	—

importance presented in Table 4.8. Interestingly, while number of servers was among the least important features in the per-page analysis, it emerged as one of the most

Table 4.6: RMSE of Random Forest per Page by Feature Selection Method

web page	Gini Importance	Permutation Importance
reddit	**0.532**	0.541
magazineluiza	**0.187**	0.189
pornhub	**0.095**	0.096
globo	**0.077**	0.079
shopee	0.110	**0.106**
olx	0.263	**0.251**
twitch	**0.346**	0.357
xvideos	0.118	**0.117**
letras	**0.063**	0.065
americanas	0.161	**0.160**
caixa	0.141	**0.140**
tiktok	**0.363**	**0.363**
amazon	0.243	**0.225**
mercadolivre	**0.114**	0.116
123movies	**0.338**	**0.338**
uol	0.295	**0.294**
spankbang	**0.276**	**0.276**
twitter	**0.229**	0.231

significant features in the e-commerce category. This suggests its crucial role in determining PLT for e-commerce pages, though it appears to be less relevant for individual page differences. The fact that over a third of the measurements originate from e-commerce pages explains why number of servers was identified as one of the most important features in the overall analysis.

In addition, for social media and news categories, the number of CSS objects and the number of JavaScript objects, respectively, were the most important features for predicting PLT. This aligns with VOGEL and SPRINGER (2022), who report that 70% of JavaScript and 90% of CSS scripts are loaded as render-blocking code, often utilized only after the page has finished rendering. Such inefficient loading represents a critical optimization opportunity. Developers targeting these categories could prioritize minimizing render-blocking CSS and JavaScript objects to improve page load times and enhance user experience.

Consistent with the per-page analysis, number of bytes and number of requests continue to be the most important features overall, further underscoring their ubiquitous influence across all categories.

4.3.2 Incorporating page category

Training one model per category was a satisfactory way of using the page category information. There are, however, other ways of making use of that information. One

Table 4.7: Selected Features by Random Forest per Page Category with Gini Importance

category	kBytes	servers	imgs.	JS	CSS	requests	dist. imgs.	pixels
e-commerce	0.145	**0.606**	—	—	—	—	0.114	—
adult content	**0.843**	—	—	—	—	0.055	0.037	—
social media	0.079	—	—	0.064	**0.713**	—	—	—
news	0.162	—	—	**0.477**	—	0.113	—	—
streaming	0.205	—	—	—	—	**0.601**	0.057	—
government	**0.773**	—	—	—	—	0.098	—	0.037

Table 4.8: Selected Features by Random Forest per Page Category with Permutation Importance

category	kBytes	servers	imgs.	JS	CSS	requests	dist. imgs.	pixels
e-commerce	0.921	0.745	—	—	—	—	**1.112**	—
adult content	**9.382**	—	—	—	2.722	1.862	—	—
social media	0.406	—	—	—	**0.904**	0.230	—	—
news	0.745	—	0.470	**2.394**	—	—	—	—
streaming	0.460	0.411	—	—	—	**1.874**	—	—
government	1.634	—	—	—	—	**5.365**	—	1.496

way would be to encode page category information into a categorical feature and add it to the general model. Another way would be to use a model that allows for different groupings of data points which, in this case, would correspond to the page categories. The second approach was adopted in this study, in the form of mixed-effects models. The first ever mixed-effect model devised was the linear mixed-effect (LME) model DUCHATEAU *et al.* (1998). The model can be summarised in the equation below:

$$\begin{aligned} Y_{ij} &= \mu + X_{ij}\beta + Z_{ij}b_i + \epsilon_{ij} && (4.1)\\ &= \mu + X_{1ij}\beta_1 + X_{2ij}\beta_2 + \cdots + X_{pij}\beta_p + Z_{ij}b_i + \epsilon_{ij} && (4.2) \end{aligned}$$

Where Y_{ij} is the response variable (PLT) for the jth navigation instance of category i, μ is the mean PLT of all the navigations, $\beta_1, \beta_2, \ldots, \beta_p$ are the fixed-effect coefficients associated to each of the p complexity metrics, $X_{1ij}, X_{2ij}, \ldots, X_{pij}$ are the complexity metric values of the jth navigation instance of category i, Z_{ij} is the random effect predictor of the jth navigation instance of category i, b_i is the random effect coefficient for category i, and ϵ_{ij} is the residual term. In matrix form we have:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}b + \epsilon \qquad (4.3)$$

This model, however, is limited by the fact that the fixed effects portion is linear. A generalization of this model would be represented by the following equation:

$$\boldsymbol{Y} = f(\boldsymbol{X}) + \boldsymbol{Z}b + \epsilon \tag{4.4}$$

In this model, the random effects portion of the modeling is linear, however the fixed effects are modeled by a general function $f()$. In the case of the mixed effect random forest (MERF) model HAJJEM *et al.* (2014), $f()$ is estimated by a random forest. This model is fit by an expectation-maximization procedure, where the following steps are followed:

1. Fix all the b_i and compute $y{*}_i = y_i - Z_i\hat{b}_i, i = 1, 2, \ldots, n$. Fit a random forest, $f(\boldsymbol{X})$, to $y*$ globally across all samples.

2. Fix $f()$, Σ_b, and Σ_ϵ and optimize to find $b{*}_i$, where Σ_b and Σ_ϵ are the covariance matrices of b and ϵ, respectively. There is a closed-form solution assuming a linear random effect and Gaussian prior.

3. Fix $f()$ and b_i and optimize to find Σ_b and Σ_e. There is a closed-form solution assuming a linear random effect and Gaussian prior.

By repeating steps 1 to 3 the model may converge, which can be calculated via the generalized log likelihood.

This model was fit to the data, resulting in an RMSE value of 0.270 and a MAE value of 0.136, making it the best-performing model so far, underscoring the importance of incorporating page categories when analyzing PLT.

4.3.3 Unsupervised Analysis

Analyzing how page complexity metrics affect PLT across different categories has proven useful, as was just shown. However, a potentially more insightful analysis could involve examining pages that share similar complexity metrics. This approach may reveal some nuances or patterns that are not apparent when comparing across broader categories. To explore this possibility, we propose an unsupervised method that groups pages based on their complexity metrics.

As a first step, the data were modeled as a three-way tensor. This was done to better understand the relationship between the different page complexity metrics, but primarily to serve as an interpretable pre-processing step for the clustering. The first mode represented the "webpage-raspberry" combination, encompassing all possible pairings of pages and Raspberry-Pis. The second mode was dedicated to complexity metrics, and the third mode to the hour of the day. For this last mode, I calculated the average value of each complexity metric for every page-raspberry pair for each hour, which formed the tensor's values. Figure 4.6 shows the tensor's representation.

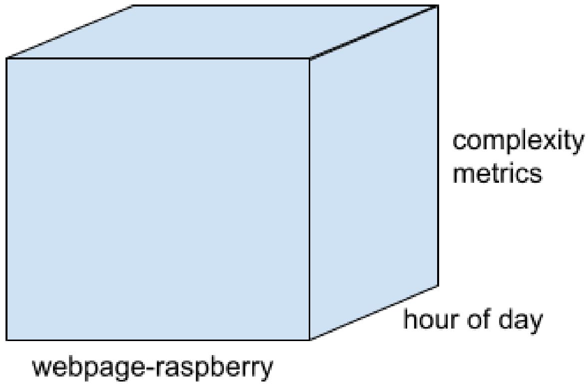

Figure 4.6: Tensor representation

Two strategies were tested for data transformation before fitting the model. First, a min-max normalization was performed, with the data being compressed into values between 0 and 1, with 0 corresponding to the lowest value and 1 to the highest, for each feature. This was then paired with a non-negative canonical-polyadic decomposition. The problem found here was that many of the features had outliers with high values, which this method of normalization is very sensitive to. Because of this, the data were normalized via standard scaling instead. Non-negative decomposition was considered; while it tends to lose precision in relation to the explained variance, its interpretation is much simpler. Nevertheless the unrestricted decomposition was not too overly complicated to interpret, which is why it was chosen for this analysis. Canonical-polyadic decomposition was then performed, using the alternating least squares algorithm. To assess the rank of the tensor, different values were tested, with split-half validation being performed and, for each candidate value, the total explained variance calculated, as shown in Figure 4.7. A rank value of 5 was chosen since the rank-5 tensor explains over 90% of the total variance in the data.

Figure 4.8 presents the factor matrix for the complexity metrics mode. The values in the matrix are called loadings, which are a way to measure how much each complexity metric contributes to each factor, allowing for a representation of the metrics in a low rank space. Considering the first three factors, all of the complexity metrics had loadings higher than 0.45. This high value suggests that every complexity metric significantly influenced at least one of these factors. Additionally, these factors were arranged in order, in which the first factor explained the most about the variations in the data, the second factor explained the next most, and so on. Since the first three factors had high loadings for all the metrics, it means they were the most significant in terms of explaining differences in the data.

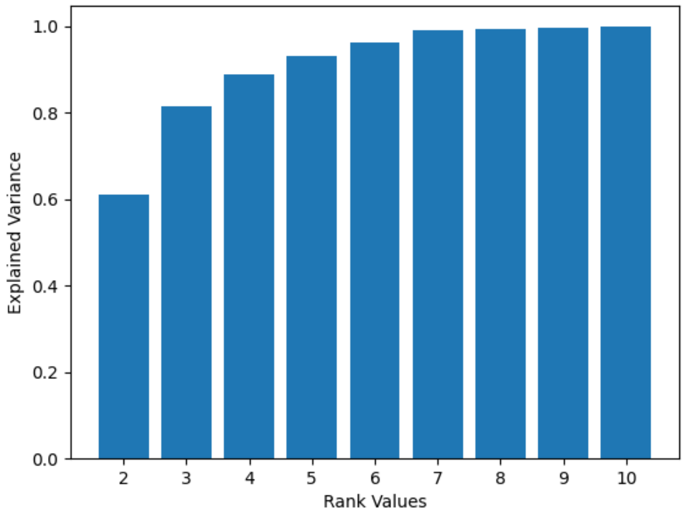

Figure 4.7: Explained variance for different ranks

Figure 4.9 displays the page-raspberry factor matrix. For clarity, only five page-raspberry combinations are depicted. As observed, the loading values for each page-raspberry pair are remarkably similar for a given factor. This similarity suggests that the patterns of page complexity metrics are consistent across different page-raspberry pairs. Given that these Raspberry Pis were connected to three distinct ISPs, this finding implies that network variations do not significantly influence these metrics, a conclusion that aligns with observations in other studies HUET *et al.* (2021). The first factor primarily accounts for the variation observed in the complexity metrics of amazon.com.br, as evidenced by the high positive loading values for the metrics: number of bytes, number of JavaScript objects, and number of distinct images (Figure 4.8). This is discussed below. In contrast, caixa.gov.br showed negative loadings for the second factor, while TikTok had positive loadings for the same factor. The second factor was mostly associated with a high number of servers, distinct images, and image objects. As expected, TikTok showed positive loadings for these metrics. Conversely, Caixa, being a government website, tended to have fewer images and servers.

Figure 4.10 shows the loadings of the hour of day mode. As can be seen, factor 1 had a distinct pattern from the others, having lower loading values from 8 to 10 in the morning. This factor had high loadings for amazon.com.br in the webpage mode and high loadings for number of JS objects, number of bytes and number of distinct images in the complexity metrics mode. Figures 4.11, 4.12, and 4.13 show the average number of bytes, JS objects, and distinct images per hour of day, respectively, for amazon.com.br, for one of the Raspberry Pis. As can be seen, all three metrics show low values between 8 and 10 am. This exact pattern was found in all but one of the Raspberry Pis, implying that amazon.com.br performs some sort of content gauging of their page according to the time of day. According to STREIT

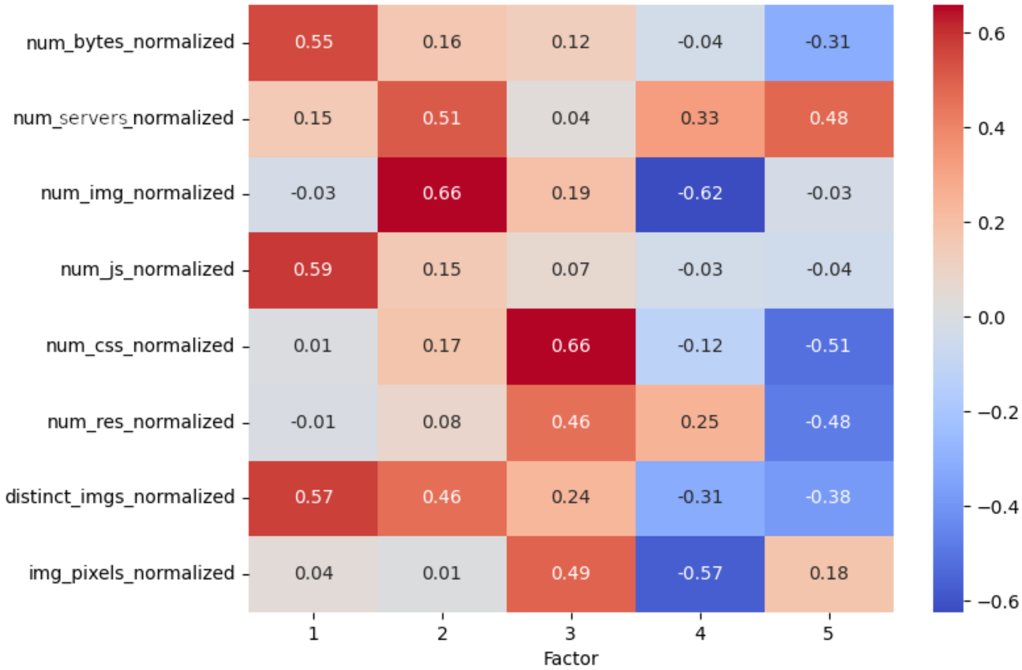

Figure 4.8: Factor Matrix for complexity metrics mode

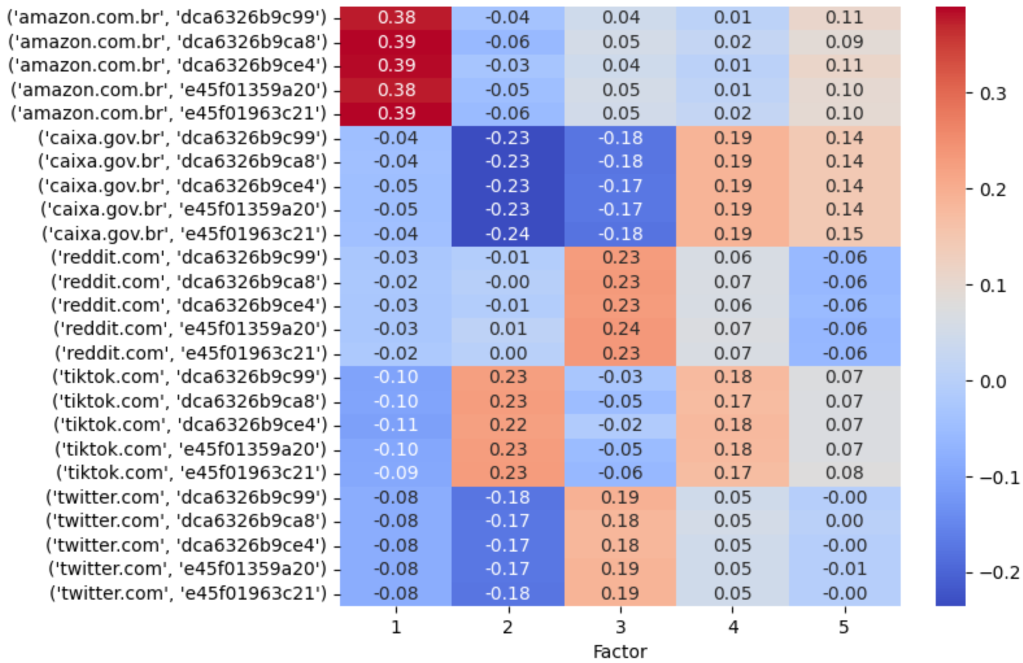

Figure 4.9: Abbreviated factor matrix for webpages mode

et al. (2023), mid-morning hours, in general, have the least internet traffic. Since there were lower than average values of JS and image objects during these hours it is likely that these originate from ad services, with Amazon opting to lower the volume of this content during hours in which the website is less accessed, in order to cut costs.

Figure 4.14 shows the average PLT values per hour of day. The lowest value is

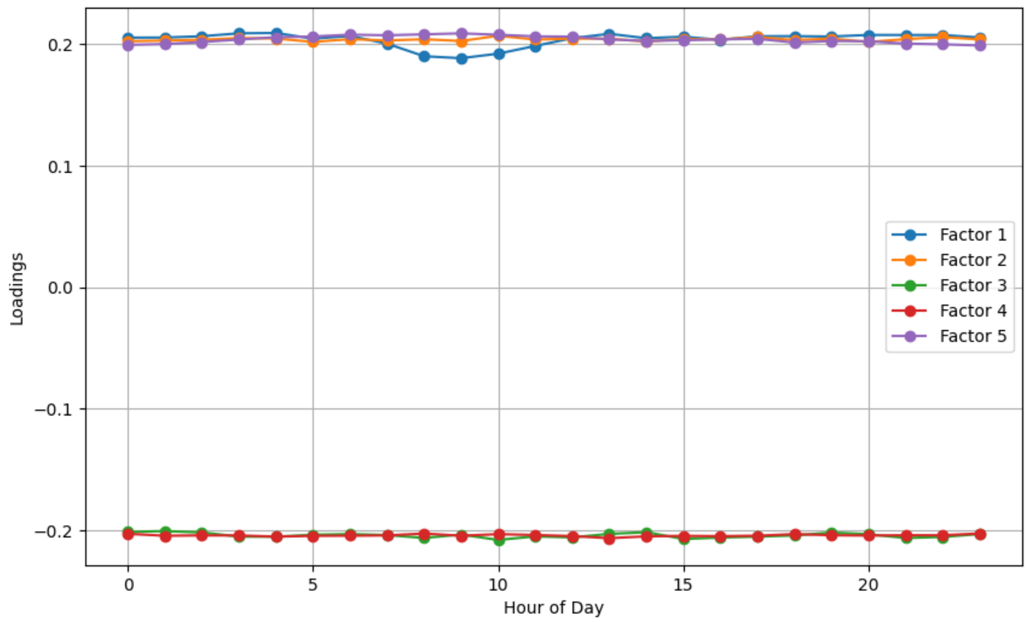

Figure 4.10: Hour of day loadings

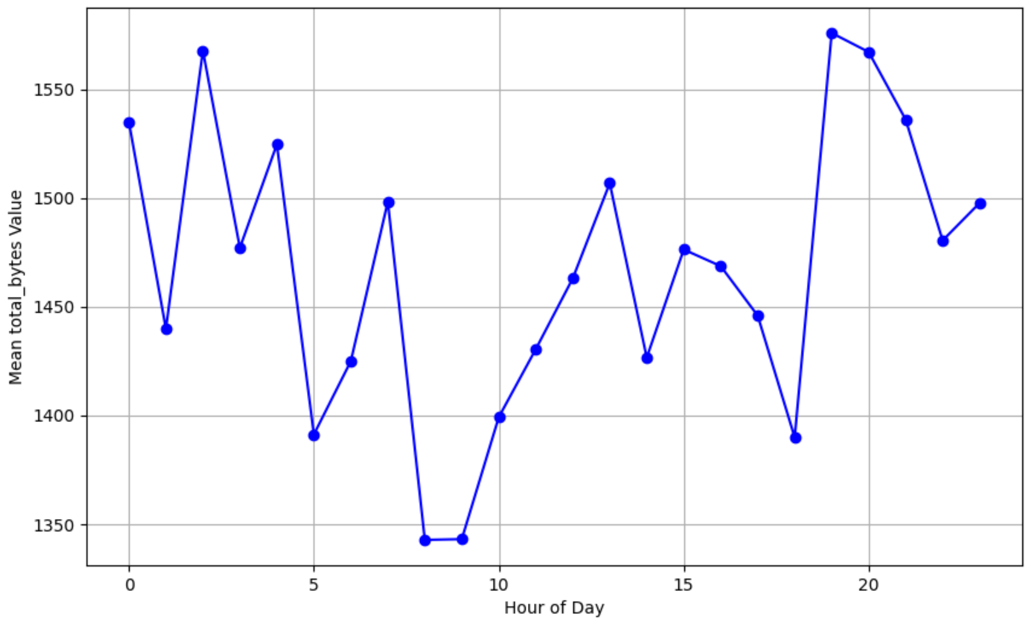

Figure 4.11: Average number of bytes per hour of day for amazon.com.br

for 9 am, while the highest is midnight. As stated before, according to STREIT *et al.* (2023), mid-morning hours, including 9 am, have the least internet traffic, while midnight is one of the hours with most internet traffic. It makes sense that hours with higher traffic usage would yield higher PLT values, since there would be higher levels of congestion in the network. The variation, though, is low, with the highest and lowest PLT values being 8555 and 8375 milliseconds, respectively. This, however, is intuitive, since most of the variation of PLT is inherent to the specific pages, not the underlying network conditions HUET *et al.* (2021).

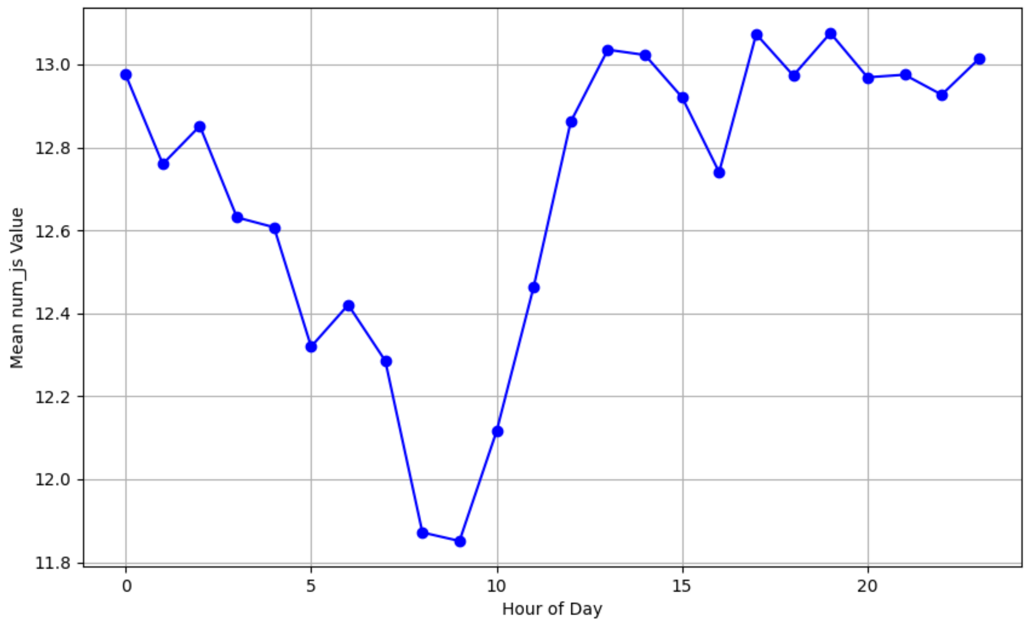

Figure 4.12: Average number of JS objects per hour of day for amazon.com.br

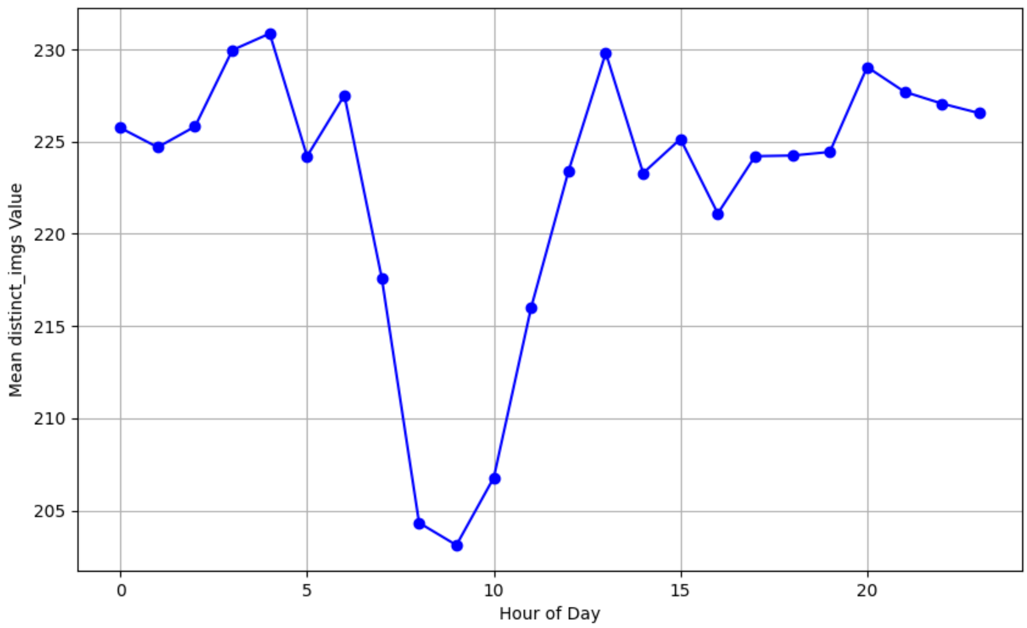

Figure 4.13: Average number of distinct images per hour of day for amazon.com.br

The data points were then clustered using the loadings from the complexity metrics mode. K-means clustering was performed with $K = i, i \in [2, 10]$. The silhouette scores were then computed for each K-value, as is shown in Figure 4.15. The elbow method was used to choose the final number of clusters to continue the analysis, which yielded a value of 4.

Figure 4.16 displays the distribution of navigations for each web page, categorized by the assigned cluster. Remarkably, almost half of all the pages had all their navigations mapped exclusively to a single cluster. Additionally, every page had at

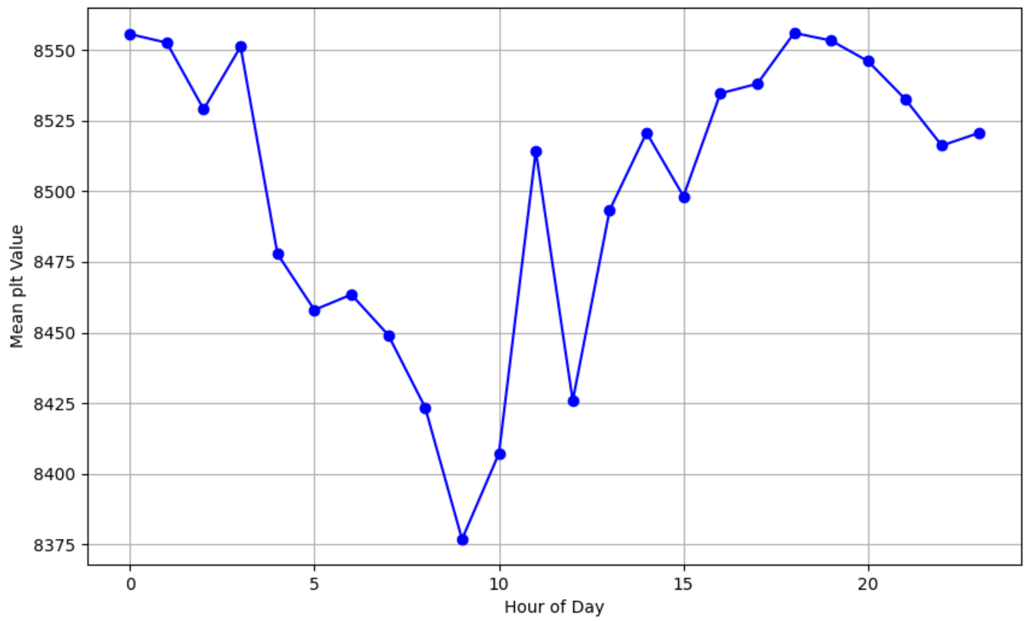

Figure 4.14: Average PLT per hour of day

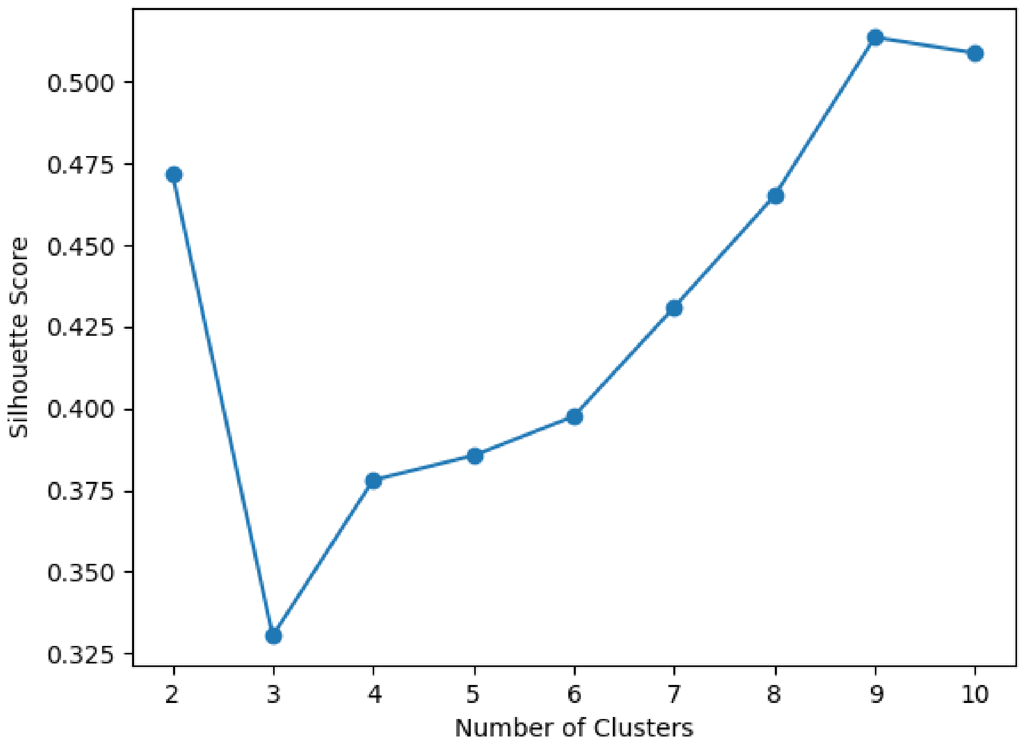

Figure 4.15: Silhouette score for different amount of clusters for K-means

least 60% of its navigations mapped to one predominant cluster. In other words, there is a remarkable correlation between the category of a web page and its corresponding cluster. Navigations to web pages featuring adult content were exclusively mapped to cluster 0, while cluster 1 predominantly grouped news and e-commerce web pages. Cluster 2 mainly grouped e-commerce and social media pages. Notably, cluster 3 was unique in containing only navigations to amazon.com.br. (Note that not all Raspberry Pis collected data from the Pornhub page, thus precluding its evaluation via unsupervised analysis.)

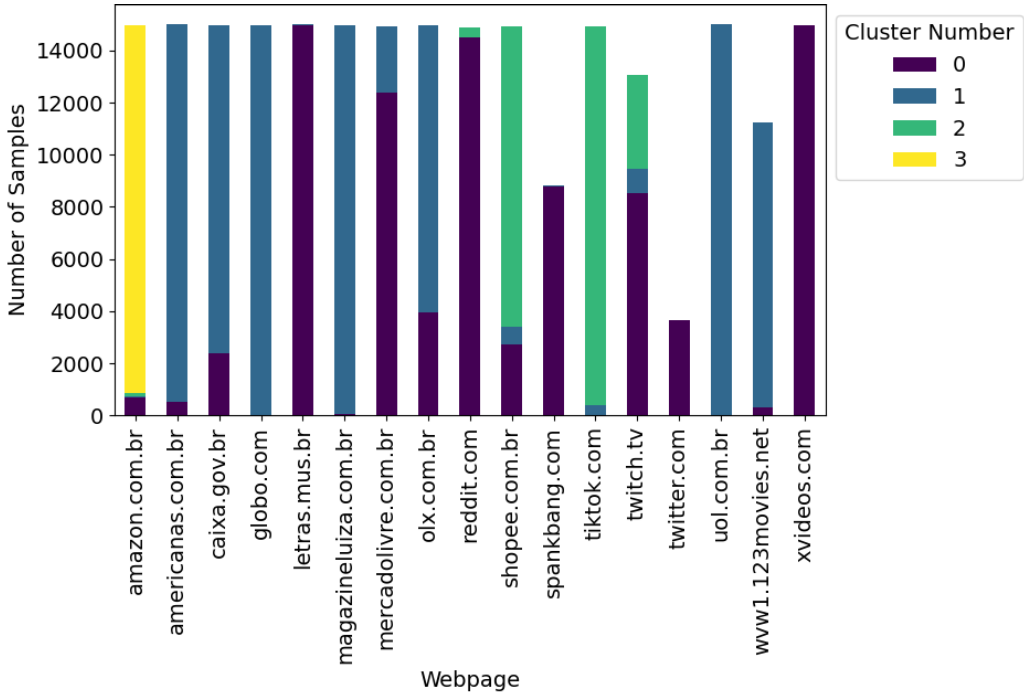

Figure 4.16: Cluster distribution per webpage

Figures 4.17 to 4.20 show the feature values for the clusters' centroids. The values are shown after being scaled via empirical standard deviation, with 0 being the empirical mean and every unit of value above or below 0 being equivalent to one standard deviation.

Cluster 0, shown in Figure 4.17, which was mainly associated with adult content webpages, presents below-average values of all complexity metrics except for number of image pixels.

Cluster 1, shown in Figure 4.18, which was comprised mainly of news and e-commerce webpages, presented nearly one standard deviation above the average number of servers and number of JS objects. This indicates a profile of pages with many ads and/or dynamic content and many different sources of content. This corroborates the findings of the previous chapter.

Cluster 2, shown in Figure 4.19, had navigation instances from the following pages: tiktok.com, twitch.tv, shopee.com.br, and very few from reddit.com, being the most diverse cluster in terms of webpage categories. The centroid of this cluster had number of CSS objects more than two standard deviations above the average and number of bytes more than one standard deviation above the average. The number of servers was nearly one standard deviation below the average. This indicates a profile of pages with few ads and lots of content, which was provided from few sources.

Cluster 3, shown in Figure 4.20, contains only amazon.com.br. The centroid has number of images nearly four standard deviations above the average and number of requests and number of distinct images nearly three standard deviations above the average. The number of servers and number of JS objects was nearly one standard deviation below average.

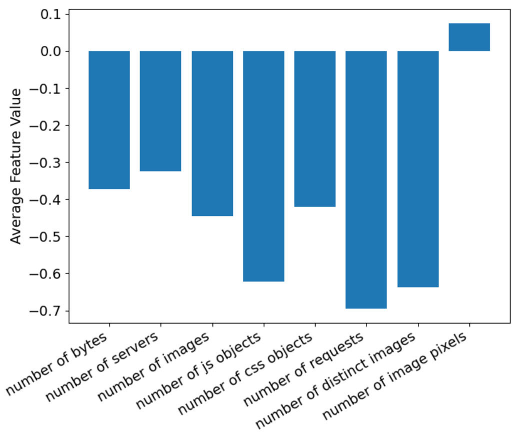

Figure 4.17: Feature values of cluster 0 centroid

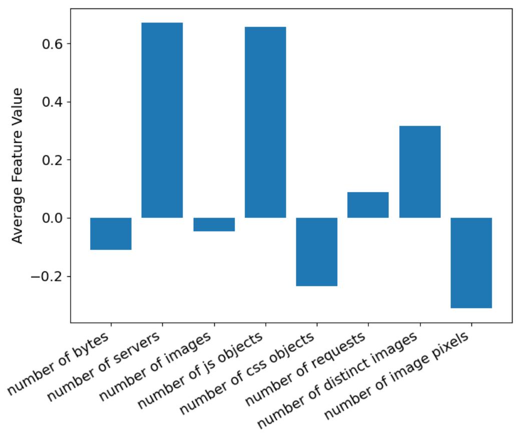

Figure 4.18: Feature values of cluster 1 centroid

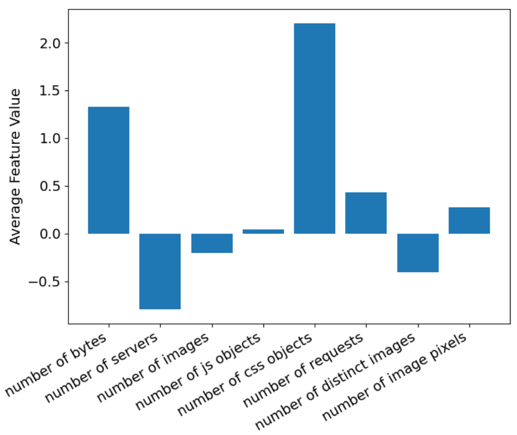

Figure 4.19: Feature values of cluster 2 centroid

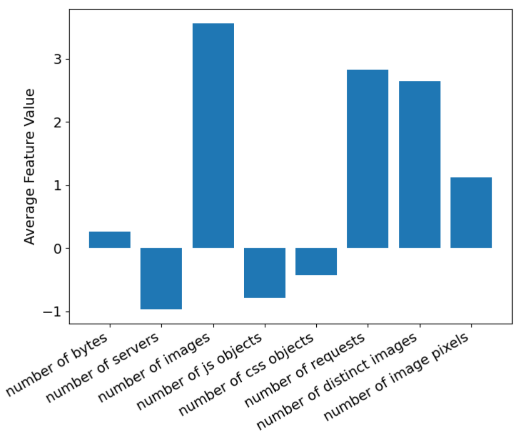

Figure 4.20: Feature values of cluster 3 centroid

Once again, Amazon was shown as a clear outlier. Upon further investigation it was found that many pages load content dynamically as the page is scrolled down. This is very common practice in e-commerce pages, which usually employ standardized content management systems (CMS) strategies. This, however, is not the case for amazon.com.br, which probably has its own in-house CMS, loading all the content in one go, allowing for smoother display of images and other types of content. To see if this accounts for the difference in profiles between Amazon and the other pages, for one week, measurements were collected differently: when a page was loaded the page was scrolled down until either the end of the page was reached or ten seconds passed. This was done because for some pages, such as twitter.com and reddit.com, content is loaded indefinitely.

Figure 4.21 shows the cluster distribution for the pages using the scrolling strategy, while Figures 4.22 to 4.25 show the cluster centroid values for the scrolling strategy. Table 4.9 shows the page migrations between clusters. The highest migration was from cluster 1 to cluster 2. While the centroids of clusters 0 and 3 remained the same, there were some small differences in clusters 1 and 2, which are highlighted in Table 4.10. H signifies a high value, above the average, while L signifies a low value, below the average. As can be seen, in the scrolling strategy, cluster 1 had number of requests and number of distinct images below average and cluster 2 had number of servers and number of JS requests above average while having number of requests and number of pixels below average. Therefore, the migration from cluster 1 to 2, in this context, indicates mainly an increase in number of bytes and number of CSS objects, which is to be expected. It can be seen that continuous scrolling for ten seconds resulted in many of twitch.tv and tiktok.com navigations being assigned to cluster 3, which was associated with high values of images, bytes, and requests. This is expected, since both are pages where not only is content loaded indefinitely, but the content also has a high number of images. Interestingly though, none of the other e-commerce pages changed to cluster 3, indicating that amazon.com.br remained an outlier in its genre.

Table 4.9: Cluster page migration fractions

Scrolling / Original	0	1	2	3
0	—	0	0.25	0
1	0.1	—	0	0
2	0.2	0.6	0	0
3	0	0.1	0.25	—

Table 4.10: Feature differences between clusters 1 and 2 for each measurement strategy

cluster	strategy	kBytes	servers	imgs.	JS	CSS	req.	dist. imgs.	pixels
1	orig.	L	H	L	H	L	**H**	**H**	L
1	scroll	L	H	L	H	L	**L**	**L**	L
2	orig.	H	**L**	L	H	H	**H**	L	**H**
2	scroll	H	**H**	L	H	H	**L**	L	**L**

Random forest was performed on each of the original clusters, with the Gini importance and permutation importance being calculated for each feature. The results are shown in Tables 4.11 and 4.12, respectively. This was the level of resolution where the selected features most differed between different feature importance methods. For both methods, though, number of bytes was the most important feature

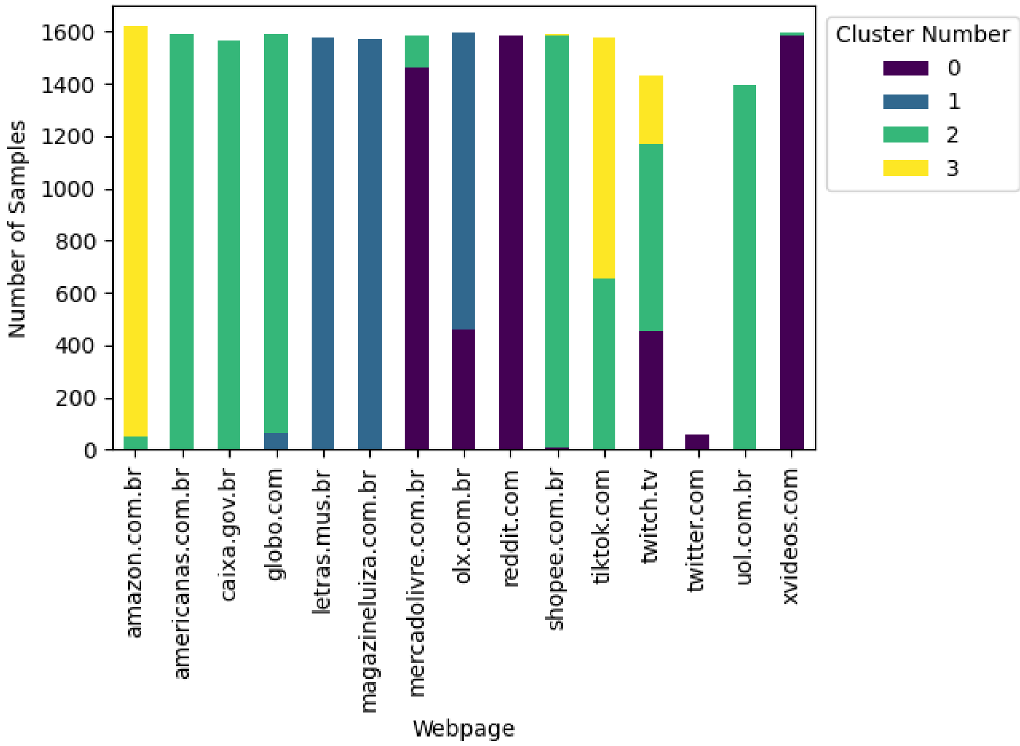

Figure 4.21: Cluster distribution per webpage with scrolling strategy

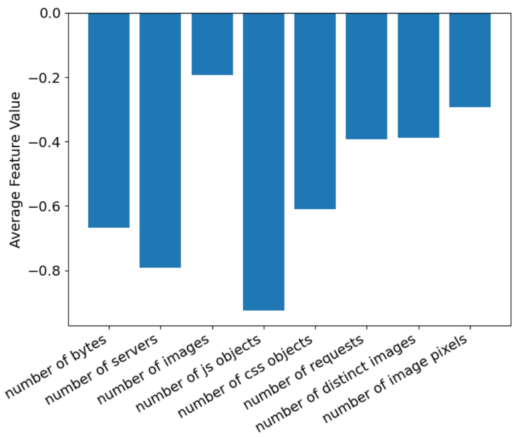

Figure 4.22: Feature values of cluster 0 centroid for scrolling strategy

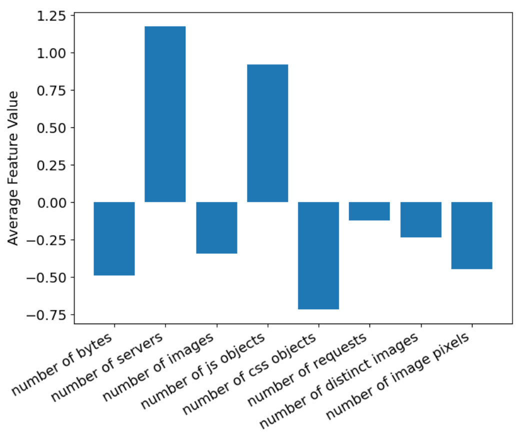

Figure 4.23: Feature values of cluster 1 centroid for scrolling strategy

for clusters 0 and 1. Number of JS objects, which was the most important feature for the news category, was also one of the most important features for cluster 1, as should be expected. Number of servers, however, which was the most important feature for e-commerce pages, was only one of the most important features for cluster 1 when employing the Gini importance method.

Interestingly, the most mixed cluster, in terms of categories, has number of servers as one of the most important features. Number of servers was also found to be the most important feature in the general analysis. This was previously understood

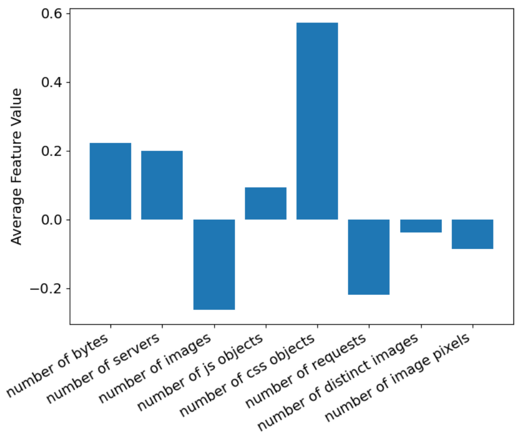

Figure 4.24: Feature values of cluster 2 centroid for scrolling strategy

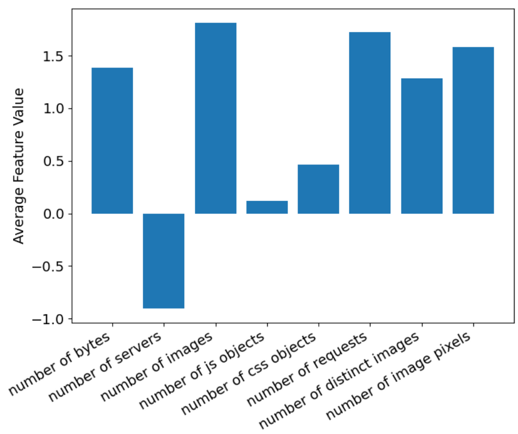

Figure 4.25: Feature values of cluster 3 centroid for scrolling strategy

to be the effect of a high number of pages from the e-commerce category, in which number of servers was found to be the most important feature, but it could in fact be the most important feature in a setting with diverse pages.

Table 4.11: Selected features by random forest per cluster with Gini importance

cluster	kBytes	servers	imgs.	JS	CSS	requests	dist. imgs.	pixels
0	**0.287**	—	—	—	0.160	—	0.111	—
1	**0.658**	0.063	—	0.131	—	—	—	—
2	0.069	**0.743**	—	—	0.063	—	—	—
3	0.139	—	—	—	—	**0.189**	0.183	—

Table 4.12: Selected features by random forest per cluster with permutation importance

cluster	kBytes	servers	imgs.	JS	CSS	requests	dist. imgs.	pixels
0	**2.174**	—	—	—	—	—	0.883	0.984
1	**1.927**	—	—	0.399	—	0.295	—	—
2	—	0.304	—	—	0.362	**0.619**	—	—
3	—	—	0.794	—	—	**0.917**	0.859	—

To compare the effectiveness of the different models, the general model, the category model, and the cluster model, with features selected via Gini importance and permutation importance, were tested on each page individually. For cluster models, the cluster assigned to each page was the one containing over 60% of that page's navigations. All data points were normalized using standard scaling, considering the entire dataset, before fitting the models.

The models were adapted to train on all of the pages except the one being analyzed. Taking the Shopee page as an example, the general model used all pages except Shopee for training, selecting the top three features for each feature selection method. This model was then exclusively tested on Shopee data. In the category model, an e-commerce model was trained (since Shopee is an e-commerce site) using all e-commerce pages except Shopee, and then tested only on Shopee data. For the cluster model, I allocated Shopee to cluster 2, as it encompasses most of its navigations. This model was trained using all data from cluster 2, excluding Shopee, and tested on all of the Shopee data, including navigations on the Shopee page that did not belong to cluster 2. The features were selected for each page/model combination, according to the training data, which, in certain cases resulted in different feature sets from those previously reported.

Table 4.13: RMSE per page for category models per feature importance method

page	category	Gini importance	permutation importance
magazineluiza	e-commerce	2.488	**1.890**
shopee	e-commerce	0.841	**0.727**
olx	e-commerce	0.870	**0.856**
americanas	e-commerce	**0.718**	1.529
amazon	e-commerce	**0.992**	1.315
mercadolivre	e-commerce	**1.342**	2.418
reddit	social media	1.173	**0.836**
tiktok	social media	1.074	**0.840**
twitter	social media	**0.896**	1.335
globo	news	0.669	**0.537**
uol	news	0.717	**0.645**
xvideos	adult	3.486	**0.416**
spankbang	adult	**0.420**	0.429
pornhub	adult	0.283	**0.198**
twitch	streaming	1.014	**0.661**
123movies	streaming	**0.698**	0.793

Table 4.13 shows the RMSE for each page, according to category models trained with each of the two feature sets. The results are absent for Caixa and Letras, as these were the only websites in the government and music categories, respectively. Although, for most pages, the model trained on the features selected via permutation importance performed better, the difference is not overwhelming, with most results being comparable.

Table 4.14 shows the RMSE for each page, according to cluster models trained with each of the two feature sets. The results are absent for cluster 3 because its data are comprised solely of navigation instances to Amazon. For this comparison both feature importance methods performed equally well, with the same number of pages having a better performance for each of the methods.

Table 4.15 shows the RMSE for each page, according to general models trained with each of the two feature sets. As was the case with the cluster analysis, the results are very balanced and, when one model has a better performance, they are

Table 4.14: RMSE per page for clusters per feature importance method

page	cluster	Gini importance	permutation importance
twitter	0	**0.811**	1.002
reddit	0	1.344	**0.994**
xvideos	0	2.920	**0.987**
spankbang	0	**0.698**	1.083
twitch	0	**0.792**	1.123
letras	0	**0.829**	1.412
mercadolivre	0	**2.140**	2.181
magazineluiza	1	1.888	**1.452**
olx	1	**0.752**	0.832
americanas	1	1.219	**0.445**
globo	1	0.868	**0.569**
uol	1	**0.831**	0.835
123movies	1	1.181	**1.081**
caixa	1	**0.578**	1.168
shopee	2	1.351	**0.979**
tiktok	2	1.429	**0.867**

Table 4.15: RMSE per page for general models per feature importance method

page	Gini importance	permutation importance
magazineluiza	**0.537**	0.947
shopee	1.242	**1.019**
olx	**0.813**	0.897
americanas	0.912	**0.863**
amazon	**1.369**	1.462
mercadolivre	**1.799**	1.920
reddit	1.641	**1.216**
tiktok	**1.734**	2.071
twitter	1.133	**0.983**
globo	0.766	**0.627**
uol	**0.942**	1.185
xvideos	1.207	**0.535**
spankbang	0.468	**0.426**
pornhub	0.324	**0.256**
twitch	**0.982**	1.000
123movies	1.416	**1.036**
caixa	1.695	**1.402**
letras	**0.954**	1.158

usually both comparable.

Table 4.16 shows a comparison between all models, in relation to RMSE for each page. It is observed that for 11 out of the 15 pages, where the category and the cluster models were evaluated, the category models outperformed both the cluster and general models. This highlights the significance of incorporating page categories in analyzing PLT. Note that, in most models, the RMSE value suggests an error in

Table 4.16: RMSE per page per model

page	category	cluster	cat. G	clus. G	gen. G	cat. P	clus. P	gen. P
magazineluiza	e-comm.	1	2.488	1.888	**0.537**	1.890	1.452	0.947
shopee	e-comm.	2	0.841	1.351	1.242	**0.727**	0.979	1.019
olx	e-comm.	1	0.870	**0.752**	0.813	0.856	0.832	0.897
americanas	e-comm.	1	0.718	1.219	0.912	1.529	**0.445**	0.863
amazon [1]	e-comm.	3	**0.992**	—	1.369	1.315	—	1.462
mercadolivre	e-comm.	0	**1.342**	2.140	1.799	2.418	2.181	1.920
reddit	social	0	1.173	1.344	1.641	**0.836**	0.994	1.216
tiktok	social	2	1.074	1.429	1.734	**0.840**	0.867	2.071
twitter	social	0	0.896	**0.811**	1.133	1.335	1.002	0.983
globo	news	1	0.669	0.868	0.766	**0.537**	0.569	0.627
uol	news	1	0.717	0.831	0.942	**0.645**	0.835	1.185
xvideos	adult	0	3.486	2.920	1.207	**0.416**	0.987	0.535
spankbang	adult	0	**0.420**	0.698	0.468	0.429	1.083	0.426
pornhub	adult	—	0.283	—	0.324	**0.198**	—	0.256
twitch	stream	0	1.014	0.792	0.982	**0.661**	1.123	1.000
123movies	stream	1	**0.698**	1.181	1.416	0.793	1.081	1.036
caixa	gov.	1	—	**0.578**	1.695	—	1.168	1.402
letras	music	0	—	**0.829**	0.954	—	1.412	1.158

estimating the PLT of less than one standard deviation. In future research, a larger number of pages could be analyzed to gain a more comprehensive understanding of category-specific trends.

Chapter 5

Conclusion

In this study, I examined the content and infrastructure of the most popular web pages in Brazil. I contrasted these findings with previous studies to understand recent trends, resulting from current page building and serving strategies. I found that e-commerce pages have changed significantly over the years in relation to number of servers and number of requests, getting closer to the profile found for news pages. Both categories now also have a higher contribution, in terms of number of bytes, in CSS and JavaScript content. In 2011, the number of bytes of JavaScript content corresponded to 39% and 34% for e-commerce and news pages, respectively. Now it is 47% and 52%. For CSS, it was 0.07% and 0.08% respectively, whereas now it is 1.4% and 1.0%. I found that the news category has the most prevalence of ad services, which is served via JavaScript files. In contrast, e-commerce pages have a much smaller presence of ads, probably using the JavaScript files mostly for dynamic pages.

I found that the median contribution of bytes which come from CDNs is almost 100% now, whereas in 2011 it was 40%. However, I also found a high correlation between use of CDNs and page rank, with higher-ranked pages having a more prevalent use. In this study only the top 100 pages were chosen, whereas in BUTKIEWICZ *et al.* (2011) the top 20000 were used. The lower-ranked pages could impact this distribution.

I performed a correlation analysis to understand the relationship between the complexity metrics. From this analysis I found a high correlation (0.77) between incoming bytes and image bytes, confirming the fact that most bytes come from images. I also found a low correlation (0.06) between number of servers and number of incoming bytes, demonstrating the high variability in number of bytes incoming from different servers. I find a high correlation between number of ad services and number of CDN servers, indicating that most ad content is hosten in CDN servers.

This study also examined the relationship between page complexity metrics and OKT using both supervised and unsupervised models. The random forest model demonstrated the best performance among the supervised models. For the unsupervised analysis, I employed a tensor decomposition approach. This method enables the representation of data in a multidimensional space, which facilitates subsequent clustering. This clustering was used to analyze how the complexity metrics affect PLT within each identified cluster.

I trained three Random Forest models: a general model (encompassing all pages), a specific page model, and a page category model, to analyze the most important features for each. The results indicate that the number of bytes is one of the top three features for the majority of models, with the only exceptions being the models for Amazon and TikTok pages.

In the page category models, the number of bytes and the number of requests emerge as the top two features for most categories. However, distinct trends appear depending on the page category. For the e-commerce category, the number of servers is the most crucial feature for determining PLT. In the news category, it is the number of JavaScript objects, likely due to the ads loaded through JavaScript. For the social media category, the number of CSS objects is most significant. These findings from the page category models align with the analysis from the specific page model, where the three most important features were identified as the number of bytes, the number of requests, and the number of distinct images.

I obtained four clusters from the unsupervised analysis. Interestingly, the number of bytes is one of the top three features for all clusters, aligning with the findings from the supervised analysis. Each cluster predominantly groups navigations to web pages within specific categories: cluster 0 exclusively contains pages with adult content; cluster 1 primarily includes news and e-commerce web pages; and cluster 2 features a combination of e-commerce and social media pages. Remarkably, cluster 3 is unique, exclusively encompassing navigations to amazon.com.br.

To assess the effectiveness of different models, I individually tested the general model, the page category model, and the cluster model on each page. The page category model outperformed both the cluster and the general models for 10 out of the 15 pages. This indicates that page category models are significant in predicting PLT. It should be noted that, in most models, the RMSE (Root Mean Squared Error) value suggests an error in estimating the PLT of less than one standard deviation.

Although these results offer significant insights, several limitations of the current analysis point to opportunities for future research. First, the study only analyzed 100 pages for the page characterization and an even smaller subset—18 pages—for the page performance analysis. Expanding the scope to include a larger dataset, encompassing pages across a broader range of page ranks beyond just the top 100, could yield more representative results. Additionally, incorporating dependency graphs into the analysis could provide a deeper understanding of inter-dependencies within page components, helping to pinpoint further optimization opportunities. These enhancements would enrich the findings and provide a more comprehensive understanding of the factors influencing PLTs.

References

AQEEL, W., CHANDRASEKARAN, B., FELDMANN, A., et al., 2020, "On Landing and Internal Web Pages: The Strange Case of Jekyll and Hyde in Web Performance Measurement", *IMC '20: Proceedings of the ACM Internet Measurement Conference*, p. 680–695.

ASRESE, A. S., ERAVUCHIRA, S. J., BAJPAI, V., et al., 2019, "Measuring Web Latency and Rendering Performance: Method, Tools, and Longitudinal Dataset", *IEEE Transactions on Network and Service Management*, v. 16, pp. 535–549.

AVRAM, C., SALEM, K., WONG, B., 2014, "Latency Amplification: Characterizing the Impact of Web Page Content on Load Times", *Lecture Notes in Computer Science*, pp. 20–25.

BISWAS, S., BICKET, J., WONG, E., et al., 2015, "Large-scale Measurements of Wireless Network Behavior", *ACM SIGCOMM Computer Communication Review*, v. 45, n. 4.

BUTKIEWICZ, M., MADHYASTHA, H. V., SEKAR, V., 2011, "Understanding website complexity: measurements, metrics, and implications", *IMC '11: Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, p. 313–328.

DUCHATEAU, L., JANSSEN, P., ROWLANDS, G. J., 1998, *Linear mixed models. An introduction with applications in veterinary research.* International Livestock Research Institute.

EGGER, S., REICHL, P., HOSSFELD, T., et al., 2012, ""Time is bandwidth"? Narrowing the gap between subjective time perception and Quality of Experience", *2012 IEEE International Conference on Communications*, pp. 1325–1330.

GAO, Q., DEY, P., AHAMMAD, P., 2017, "Perceived Performance of Top Retail Webpages In the Wild: Insights from Large-scale Crowdsourcing of Above-the-Fold QoE", *Internet QoE '17: Proceedings of the Workshop on QoE-based Analysis and Management of Data Communication*, p. 13–18.

GIGASPACES, July, 2023. "Amazon Found Every 100ms of Latency Cost them 1% in Sales". Available at: <https://www.gigaspaces.com/blog/amazon-found-every-100ms-of-latency-cost-them-1-in-sales>. https://www.gigaspaces.com/blog/amazon-found-every-100ms-of-latency-cost-them-1-in-sales, Accessed: 2024-01-12.

GLOAGEN, L., 2022. "Dynamic Serving vs. Responsive Web Design". Available at: <https://www.spiria.com/en/blog/mobile-development/dynamic-serving-vs-responsive-web-design/>.

GOOGLE, April, 2010. "Using site speed in web search ranking". Available at: <https://developers.google.com/search/blog/2010/04/using-site-speed-in-web-search-ranking>. https://developers.google.com/search/blog/2010/04/using-site-speed-in-web-search-ranking, Accessed: 2024-01-12.

HAJJEM, A., BELLAVANCE, F., LAROCQUE, D., 2014, "Mixed-effects random forest for clustered data", *Journal of Statistical Computation and Simulation*, pp. 1313–1328.

HORA, D. D., ASRESE, A. S., CHRISTOPHIDES, V., et al., 2018, "Narrowing the gap between QoS metrics and Web QoE using Above-the-fold metrics", *PAM 2018 - International Conference on Passive and Active Network Measurement*, pp. 1–13.

HOSSFELD, T., METZGER, F., ROSSI, D., 2018, "Speed Index: Relating the Industrial Standard for User Perceived Web Performance to web QoE", *2018 Tenth International Conference on Quality of Multimedia Experience*, pp. 1–6.

HUET, A., SAVERIMOUTOU, A., HOUIDI, Z. B., et al., 2021, "Deployable Models for Approximating Web QoE Metrics From Encrypted Traffic", *IEEE Transactions on Network and Service Management*, v. 13, pp. 3336–3352.

JAHROMI, H. Z., DELANEY, D. T., HINES, A., 2018, "How Crisp is the Crease? A Subjective Study on Web Browsing Perception of Above-The-Fold", *2020 6th IEEE Conference on Network Softwarization*, pp. 43–50.

MOLNAR, C., 2019, *Interpretable Machine Learning*. Lulu.com.

NEOTERIC, 2016. "Single-page application vs. multiple-page application". Available at: <https://medium.com/@NeotericEU/single-page-application-vs-multiple-page-application-2591588efe58>.

SALUTARI, F., HORA, D. D., DUBUC, G., et al., 2019, "A Large-scale Study of Wikipedia Users' Quality of Experience", *WWW '19: The World Wide Web Conference*, p. 3194–3200.

SAVERIMOUTOU, A., MATHIEU, B., VATON, S., 2019, "A 6-month Analysis of Factors Impacting Web Browsing Quality for QoE Prediction", *Computer Networks*, v. 164.

STREIT, A. G., RIBEIRO, M., LEÃO, R. M. M., et al., 2023, "Residential Traffic Profiles: What Could Be Learned from Lightweight Measures During the Pandemic?" *SSRN*.

VOGEL, L., SPRINGER, T., 2022, "An In-Depth Analysis of Web Page Structure and Efficiency with Focus on Optimization Potential for Initial Page Load", *International Conference on Web Engineering*, pp. 101–116.

WIEGAND, April, 2022. "Site Speed is (Still) Impacting Your Conversion Rate". https://www.portent.com/blog/analytics/research-site-speed-hurting-everyones-revenue.htm, Accessed: 2024-01-12.