What do Computing Interns Discuss Online? An Empirical Analysis of Reddit Posts

Saheed Popoola
School of Information Technology
University of Cincinnati
USA
saheed.popoola@uc.edu

Ashwitha Vollem
School of Information Technology
University of Cincinnati
USA
vollemaa@mail.uc.edu

Isaac Kofi Nti
School of Information Technology
University of Cincinnati
USA
ntiik@ucmail.uc.edu

Abstract

Internships are the most common way for students to gain practical real world experience, and they have become a major part of most computing curricula because they help match student abilities or expectations with the demand of the workforce. The internship experience for students vary due to diverse factors such misconceptions about industrial realities, levels of preparation, networking opportunities and so on. Existing research on internship process often used survey or interviews to collate student's opinion of the internship experience. Unfortunately, this approach may not capture the diversity of intern experiences across different geographical regions due to limited number of participants. This paper provides insights into interns' discussion on social media. We extracted 143,912 online Reddit posts related to computing internships out of 921,845 posts containing the root word "intern", and then used topic modeling to unravel the common themes in the discussions. Next, we applied sentiment analysis techniques to understand the feelings expressed by students in the Reddit posts. The results show that computing interns generally express a positive sentiment, and the discussions were mostly related to academics, school admissions, professional career, entertainment activities, and social interactions.

CCS Concepts

• **Social and professional topics → Computing education**.

Keywords

interns, online forums, topic modeling, sentiment analysis

ACM Reference Format:
Saheed Popoola, Ashwitha Vollem, and Isaac Kofi Nti. 2025. What do Computing Interns Discuss Online? An Empirical Analysis of Reddit Posts. In *26th Annual ACM Conference on Cybersecurity and Information Technology Education (ACM SIGCITE 2025), November 6–8, 2025, Sacramento, CA, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3769694.3771103

1 Introduction

Traditional approaches to computing education often limit students' exposure and engagement with real world projects, thereby failing to fully harness their potential and creativity [10, 11]. The industry has often noted that fresh graduates have limited exposure to real world projects and are often ill-equipped to handle industrial projects [8, 27, 31, 32]. Against this backdrop, internships offer a valuable opportunity for students to gain practical, real world experience.

An internship is a structured program where students are trained in the industry to work on real world projects [14]. The internship often occurs when the school is on break, typically in the summer. A number of research has been conducted to understand the impact of internships on students [16, 23, 26], the factors that contribute to successful or unsuccessful internship experiences [9, 15], and the challenges students encounter during internships [25]. However, these studies often relied on surveys or questionnaires that may involve limited number of participants and fail to capture the diversity of intern experiences across different geographical regions.

Online social media platforms such as Reddit provide a space for users to share their thoughts on various topics. A detailed analysis of how people behave in online forums may offer deeper insights into the fundamental mechanisms by which collective thinking emerges in a group of individuals. In 2022 alone, Reddit had 277 million posts with 1.43 billion comments[1]. Furthermore, Reddit data is relatively open compared to Twitter or Facebook [4]. This abundance of publicly available data makes Reddit a popular resource for many academic research [24, 30].

This paper contributes to the use of social media platforms for educational purposes and extends the literature on internships to include the opinions of a wider group of students via Reddit posts. The paper analyzes the discussions of computing interns on the Reddit platform. We extracted over 140,000 posts related to computing internships and applied topic modeling techniques to gain insight into the common themes in the intern discussions. We then applied sentiment analysis on the post to extract the aggregate emotions in the posts. This paper makes the following important contributions.

(1) The paper introduces a use case on the use of Reddit for educational purposes. The paper also discusses how we filtered the noise in the collected data.
(2) A publicly available dataset[2] on computing internships. The dataset contains 143,912 posts (from June 2009 to December 2023), and the Python scripts used for data collection and data analysis.
(3) A description of the common themes in computing internship discussions. This offers a deeper understanding of the

ACM SIGCITE 2025, Sacramento, CA, USA

ACM ISBN 979-8-4007-2240-0/2025/11
https://doi.org/10.1145/3769694.3771103

[1] https://www.statista.com/topics/5672/reddit/#editorsPicks
[2] https://github.com/compedutech/Intern-Forum

major factors that contributes to a successful or unsuccessful internship experience.

(4) The study highlights the overall sentiment expressed by interns, shedding light on their collective experiences.

The contributions of this paper can inform further research investigations into how to ensure students can have a successful internship experience and satisfy the main objective of the internship program i.e., for students to gain practical, real world experience.

2 Literature Review

Internship helps to bridge the gap between the theoretical knowledge students gain in formal academic settings and its practical applications in the real world. Thus, internships enhance the career prospects, personal growth and intellectual curiosity of students [28, 38]. The experience gained at internships also builds confidence and helps students to better understand their career preferences [2, 25]. Existing research have shown that internships can help students to acquire practical knowledge, understand industry-work environment and secure future employment opportunities [13, 15]. However, there is still limited knowledge about why some students chose not to intern before they begin fulltime work, even though previous studies have shown that 40% of students prefer to intern at least once before they graduate [16]. This highlights the need to understand student perception of internships. Furthermore, internships play a significant role in the job recruiting process because it allows companies to evaluate student abilities over an extended period [14, 35].

Similarly, internships provide opportunities for students to develop professionalism and understand workplace issues in the field of computing [19]. The experience gained through internships often help students to perform better in future. Binder et al. [5] conducted a study on about 15,000 students to examine the impact of internships on the students' academic performance. Their results show that students who have completed an internship program have better academic performance regardless of the students' natural talent. This finding is consistent with the results reported by Jamie et al. [13]. Blicblau et al. [7] also reported that students who have done internship for longer time have better academic performance compared to students who have done internship for shorter time.

Although computing internships benefit students in many ways, they also introduce various challenges. For example, students must adapt to the work culture of the company, act professionally, and learn how to apply their theoretical knowledge practically. Therefore, the internship experience maybe overwhelming for many students. To mitigate some of these challenges, Sweetser et al. [36] came up with design, implementation and evaluation guidelines for computing internships that benefit both the employer and student.

Data mining has become an important tool to analyze data, gain better insights, and improve learning experiences within the field of computing education [1]. The application of data mining techniques to analyze online discussions can provide a better understanding of students' concerns and challenges. Techniques such as sentiment analysis can help to gauge students' perceptions and opinions on various programs, including computing internships [34]. These perceptions can generate actionable insights on how to improve these programs.

The hands-on experience students gain from internships will significantly improve students' readiness for their future career and promote their professional growth. Similarly, data mining techniques can be leveraged to analyze students' data and improve learning outcomes [3]. Therefore, by applying data mining techniques on internship related data, researchers can have better understanding of the experiences of computing interns. These insights can enable students, universities and employers to collaborate and design better internship programs. This paper advances the literature by using social media to explore major discussion themes and sentiment of students in internships.

3 Methodology

Online forums provide a platform for users to post questions, share experience, and reply to a wide range of topics. A comprehensive analysis of the discussions related to a specific concept such as internships, can offer valuable insights into the patterns of thought and emergence of collective opinions within a group of individuals. Such analysis can help identify the core challenges interns face, the expectations students have during an internship, and the factors that contribute to a successful or unsuccessful internship experience.

The objective of this study is to gain a deeper understanding of the online discussions of computing interns. To achieve this, the study is guided by the following three research questions.

- **RQ1**: What are the prevalent themes in the discussions related to computing internship?
- **RQ2**: What sentiments are expressed in discussions about computing internships?
- **RQ3**: What are the major domains covered in the discussions related to computing internships?

3.1 Data Collection

We downloaded the Reddit data dump[3] provided by Pushshift [4] that contains data from June 2009 to December 2023. From this dataset, we extracted posts that contain four key words: "intern", "interns", "internship", and "internships". This search process via keywords mirrors the methodology used in Systematic Literature Reviews[17] where initial relevant papers are extracted based on a set of keywords and boolean operators. The initial search results produced 921,845 posts. However, this initial dataset included significant noise that are not related to computing internship. To address this, we used the HuggingFace implementation[4] of the Facebook Bart model [20] to classify posts in the initial dataset as computing or non-computing. At the end of the classification process, 308,841 posts were classified as computing and 613,004 posts were classified as non-computing. We evaluated the model's performance by randomly selecting over 100 posts that were then independently and manually classified by the authors to establish a ground truth. We only used posts where are all the authors agreed on the classification to evaluate the Bart model. The authors used some inclusion and exclusion criteria to determine if a post is related to computing or not. The inclusion criteria for computing post are.

[3] https://academictorrents.com/details/9c263fc85366c1ef8f5bb9da0203f4c8c8db75f4
[4] https://huggingface.co/facebook/bart-large

- Authored by a computing major. This examines if the author of the post has a background in a computing field such as information technology, computer science, computer engineering, software engineering, etc.
- Relevant companies. Posts that mention companies that are involved in computing-related activities.
- Technical terms. Posts that contain technical terms or industry-specific language related to computing field.
- Discussions of computing technologies. Posts that address any computing technologies or methodologies.
- Programming contents. Posts with code snippets or programming tutorials.
- Posts that cite or reference any articles, books, or research papers related to computing topics.

The following exclusion criteria was used to classify posts as non computing.

- Non-compliance with the inclusion criteria. The post did not meet any of the inclusion criteria for computing.
- Non-technical work. Posts that share professional and personal experiences related to non-computing such as marketing, finance etc.
- Discussion of non-computing fields. Posts addressing non-computing topics.
- Discussion of non-technical challenges. Posts discussing non-technical experiences outside the scope of computing like interpersonal relationships, travel experience

The comparison of the Bart's classification against the authors' classification shows that there were a lot of false positives in the Bart's prediction. 65% of the posts classified as computing by the Bart model were actually computing related when compared with the posts rated by humans. To enhance the quality of the dataset used for the study, we further filtered the Bart-classified computing posts by their subreddit information. We selected the top 200 subreddits with the highest number of posts for the 308,841 posts classified as computing related by the Bart model. We manually analyzed each subreddit description and filter out subreddits that are not related to computing such as subreddits dedicated to relationships, subreddits reserved for mature audience, or subreddits dedicated to a non-computing discipline such as accounting. However, we retained generic subreddits like immigration and school admissions because a lot of computing interns also shared their internship experience in these subreddits. At the end of the process, the dataset contained 143,912 posts from 159 subreddits (top 200 subreddits minus non-computing subreddits). We believe the combination of Bart classification and manual subreddit filtering produced a quality dataset for the analysis conducted in this study. Figure 1 provides a graphical summary of the data collection process.

The complete dataset including the original unclassified posts, unfiltered classified post, filtered classified posts, the scripting codes used for filtering and classification, and the manually classified posts by the authors are publicly available in a GitHub repository[5] For the rest of this study, we focused on the 143,912 posts that was filtered based on their subreddit. We believe this approach successfully captures a substantial portion of discussions related to computing internships on Reddit.

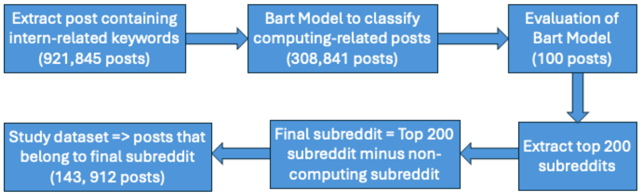

Extract post containing intern-related keywords (921,845 posts)
Bart Model to classify computing-related posts (308,841 posts)
Evaluation of Bart Model (100 posts)
Extract top 200 subreddits
Final subreddit = Top 200 subreddit minus non-computing subreddit
Study dataset => posts that belong to final subreddit (143, 912 posts)

Figure 1: The data collection process

3.2 Study Design

For the study, we used topic modeling and sentiment analysis techniques to answer the research questions. The following subsections provide a detailed description of the study design.

3.2.1 Overall Pipeline of Topic Modeling. Topic modeling is a process that uses unsupervised machine learning techniques to identify topics within a text without human input. This method is commonly used to uncover underlying thematic structures in the content of a text [22]. The process typically involves three main steps: preprocessing, topic modeling, and results interpretation. First, we refined the texts by removing meaningless characters (such as '[', punctuations, etc) and common stop words. We then used stemming techniques to reduce words to their base form (e.g., changing 'worked' to 'work'). Secondly, we used a topic modeling algorithm called Latent Dirichlet Allocation (LDA) [6] from Gensim Python package [33] to identify topics or themes within the text. The LDA algorithm generated a set of keywords associated with each topic that defined the major themes in the text. Thirdly, we applied open coding techniques to interpret the results by manually assigning descriptive labels to the topics based on the keywords. The labels were assigned after reading sample posts with high probabilities for a specific topic to understand the context better. The main goal of the labeling process was to summarize the prevailing themes associated with each identified topic.

3.2.2 Sentiment Analysis. Sentiment analysis is a branch of Natural Language Processing (NLP) that is used to determine the emotional tone conveyed in text [29]. This analysis helps to understand the sentiment expressed within the text and they are typically classified as positive, negative, or neutral. We used two popular Python packages TextBlob [21] and Vader [12], to extract the sentiments in interns' posts. Both tools return a sentiment analysis score ranging from -1 to 1, where -1 is indicates extremely negative sentiment and 1 represents extremely positive sentiment. We followed the guidelines outlined by Oyebode & Orji [29] to categorize posts based on the analysis scores. Table 1 summarizes the rules used for classifying the post based on the output from Textblob and Vader. The final classification rule adopted is that a post is classified as positive if both Textblob and Vader returned a positive classification, while the post is classified as negative if both tools returned a negative classification; otherwise, the post is classified as neutral. We used a range of -0.05 to +0.05 for neutral classification in Vader because this generates a very high accuracy according to Huto & Gilbert [12]. The combined approach of Vader and Textblob ensured a more robust and accurate sentiment classification.

[5] https://github.com/compedutech/Intern-Forum

Classification	Textblob	Vader
Positive	> 0	> 0.5
Negative	< 0	< -0.5
Neutral	0	≥ −0.5 && ≤ 0.5

Table 1: Classification rules based on analysis scores

4 Results

This section presents the results of the data analysis discussed in the previous section. The results have been grouped based on the research questions introduced in Section 3. To answer these research questions, we have used the topic modeling and sentiment analysis techniques described in Section 3.2.

4.1 RQ1: What are the prevalent theme in the discussions related to computing internship

We used to coherence score [37] of the LDA model to determine the optimum number of topics by comparing the coherence scores across models with 2 to 20 topics. The results shows that intern discussions are centered around five main themes. Sample posts containing the keywords were extracted, read, and manually analyzed to better understand the context and interpret the topic modeling results. The identified themes are academics/admissions, professional career including job applications, entertainment activities, emotions/social interactions, and the core internship experience (mostly as software engineer). Each of these themes are discussed below.

(1) Entertainment. This cluster shows that interns often discuss about entertainment, including gaming options. This is also particularly important since many of the interns are new to the industry work environment. Furthermore, some of the post also includes suggestions for games that fosters team bonding.

(2) Core internship experience. This cluster includes discussions on the student's experience during the internship at the workplace. Many posts covers almost all aspects of the internship process including the interview process, navigating multiple internship offers, the type of job assigned, project they worked on, and what they gained from the internship experience. The cluster also showed that majority of the positions were related to software development.

(3) Social interactions. This cluster deals with post on social interactions. Many of the interns usually move to a new city for their internship and they often enquire about how to make friends and places to visit. Some of the posts also covers the social interactions among the employees in the work place.

(4) Academics and admissions. This cluster covers discussion related to their academic programs and how an internship experience may contribute to a successful admission to an academic program.

(5) Professional career. This cluster contains discussions centered around career help, particularly for jobs and internships in the field of computer science. Many posts likely involve people seeking advice about job searches, navigating careers in tech, and new internship opportunities.

Topics	Key words
Topic 1 - Entertainment	itunes, vudu, itunesport, movie, google, point, amp, war, play, season, star, anywhere, unrated, man, collect, day, split, playport, port, dark game
Topic 2 - Core internship experience	company work, experience, job, interview, get, offer, intern, engineer, data, year, like, apply, develop, software, position, learn, know, project,look
Topic 3 - Social interactions	like, one, get, time, work, know, want, back, day, thing, people, could, look, make, said, think, try, got, ask, year
Topic 4 - Academics and admissions	school, year, university, research, work, student, science, college, program, one, gpa, class, club, computer, intern, major, also, get, apply, math
Topic 5 - Professional career	get, job, work, year, like, time, want, feel, know, school, really, start, graduate, take, college, also, make, one, think, experience

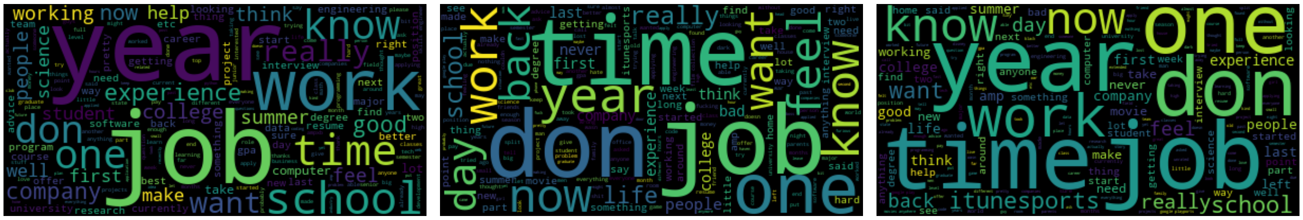
Table 2: Topic and keywords

This result suggests that interns view the internship experience as a vital part of the school admission or job application process. The analysis highlights the interconnection between interns' personal and professional lives as they still had to engage in social interactions alongside their technical responsibilities. Finally, the results may also indicate that majority of the computing interns work in a software engineering role during their internship. Table 2 shows the keywords associated with each of these five topics.

4.2 RQ2: What sentiments are expressed in discussions related to computing internships

Although the discussion topics in RQ1 is important for understanding prevailing themes, it does not capture how students feel about these themes. For example, some students expressed frustration that their academic program required them to have an internship training but the students have not been able to secure one. From the context of the post, it appears that the frustration was directed towards the internship requirement and not their inability to secure an internship position. This highlights the importance of understanding the sentiments that underlies these discussions.

The results of the sentiment analysis process described in Section 3.2.2 show that out of 143,912 posts, 104,634 (72.7%) were classified as positive, 7,939 (5.5%) were classified as negative, and 31,339 (21.8%) were classified as neutral. This result suggests that majority of the interns were satisfied with their internship experience. Figure 2 presents the word cloud for each of the sentiments. The figure show that the words "year", "job", and "work" are prominent for positive sentiments. This may indicate overall satisfaction with the time of the internship and the assigned tasks. The words "feel", "know" and "don" (stemmatized don't) were prominent with negative sentiment. This may indicate the technical and emotional challenges the interns experience.



(a) Positive Sentiment **(b) Negative Sentiment** **(c) Neutral Sentiment**

Figure 2: Word cloud for positive, negative, and neutral sentiments

4.3 RQ3: What are the major domains in the discussions related to computing internships

The Reddit platform is composed of subreddits that targets a specific topic or domain. The description of each subreddit typically outlines the expected subject area for posts created in the subreddit. Furthermore, every post on Reddit is associated with exactly one subreddit, thereby providing a natural categorization mechanism for the post. We analyzed the subreddits associated with the dataset to understand the major subject areas that computing interns are interested in. This analysis of the subreddits in the collected data provided valuable insights into the intended subject matter of the posts. The final dataset used for this analysis contains 159 subreddits that covers a wide range of topics. The most popular subreddit was "cscareerquestions" with 30,209 posts or about 20% of the total post. According to the subreddit description, cscareerquestions is *for those who plan to enter or already in the computer science field to help them navigate the challenges of the industry and share strategies to be successful.* The next top subreddits are csMajors (12,017 posts), chanceme (chances of acceptance to college) (5,117 posts), ITCareerQuestions (5,026 posts), jobs (4,648 posts), careerguidance (4,033 posts), and learnprogramming (3,993 posts). We manually reviewed the description of each subreddit in order to identify the subject niche for that subreddit. We then grouped subreddits with similar subject matters into major domains. The results of the analysis reveal eight major domains described below.

(1) Academics (including admissions). This domain encompasses subreddits related to academics including admissions, classwork, university life, and so on. Examples of subreddits in this domain are gradadmissions, ApplytoCollege, CSEducation and university-specific subreddits such as Cornell, UTAustin, and Purdue. This domain contains the highest number of subreddits.

(2) Career. This domain covers all the subreddits related to career including how to get jobs or internships, career pathways, job experience, resume building, work-life balance, and so on. Example of subreddits in this domain includes cscareerquestions, jobs, careerguidance, techjobs, etc. This domain has the highest number of posts.

(3) Data science. This domain focuses on data science and machine learning. This domain has the fewest subreddits and posts. Sample subreddits in this domain are learnmachinelearning, dataengineering, and datascience.

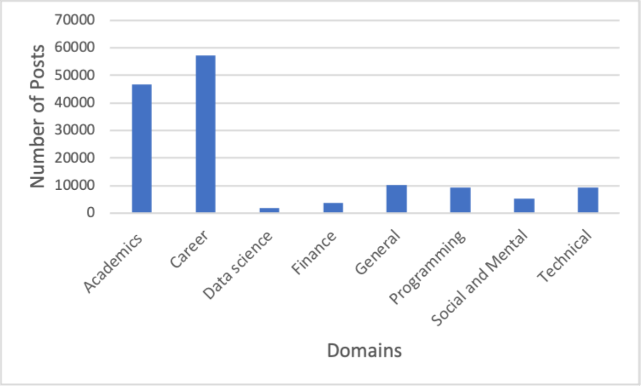


Figure 3: Number of posts in each domain

(4) Finance. This domain covers all subreddits related to financial advice or financial companies. Examples of subreddits in this domain include FinancialCareers, Big4, and quant.

(5) General. This domain includes subreddits that are generic and not specific to a niche. Some of the subreddits are specific to specific countries. Examples of subreddits in this domain are Advice, AskReddit, LifeAdvice, India, and askSingapore.

(6) Programming and software engineering. This domain targets programming, software development, and web development. Although a significant number of software engineering related posts are present in all other domain, we included subreddits that targets only software engineering related activities in this domain. Examples of subreddits include learnprogramming, learnpython, typescript, and SoftwareEngineering.

(7) Mental Health and social interactions. This domain focuses on mental health and social interactions. The domain includes subreddits such as mentalhealth, socialanxiety, and therapists.

(8) Technical competence. This subreddit covers all other technical subreddits (related to engineering or computer science) except for data science and software engineering. Examples of subreddits in this domain include sysadmin, InformationTechnology,techsupport, and AsksEngineers.

There were also a number of subreddits that belong to more than one domain. In this scenario, we assigned the subreddit to the more dominant domain. For example, developersIndia, a subreddit for software developers in India, is classified under the programming and software engineering domain instead of the general domain. Figure 3 presents an overview of the number of posts in each domain. It can be seen that majority of the posts in the domains are related to career, seeking technical knowledge, academics, and admissions. This is also consistent with the RQ1 results reported earlier in Section 4.1.

5 Threats to Validity

This section discusses the threats that may affect the validity of our research. It is possible that some of the posts in the final dataset may have been incorrectly classified as computing-related posts. It is also possible that some of the posts may be assigned incorrectly to a topic during the LDA topic modeling process. However, there is no perfect solution in the data collection or topic modeling process that guarantees accurate classification for all documents. Hence, while the manual verification process indicates that the Reddit mining approach used in this paper produced reliable results, some posts may have been misclassified. Furthermore, this study covers the aggregate posts from 2009 to 2023, and we did not separate the posts by year. Hence, it is possible that events in some years (e.g., COVID) might greatly influence the sentiments for these years. This is also true for other features that were not considered in this study such as the demographics of the post author. We plan to address this limitation in future work.

Finally, the study relied on only one source (i.e., Reddit) for the data used in the study. While Reddit provides a rich and diverse set of posts, other online discussion forums, such as StackOverflow[6] could offer additional insights into intern discussions. This threat may affect the generalization of our findings. For example, StackOverflow is more programming-oriented and it is possible that the topics embedded in StackOverflow posts may be more technical. To address this threat, we plan to include data from StackOverflow in our future research work. The expansion of the dataset to include multiple sources will help validate the findings and provide a more comprehensive understanding of the topics and sentiments expressed by computing interns across different online communities.

6 Discussion

The results of this study provided some insights into the experience of computing interns. We propose some recommendations based on the results that may be used to enhance the computing curriculum.

(1) Emphasis on career preparation and guidance. The RQ1 and RQ3 results show that students are interested in career-related discussions. A deeper look at some of the posts show that students often seek career advice such as guidance on navigating internship applications and transition to industry roles. To address this, computing curricula should include career counseling and professional development modules. These could cover resume building, interview preparation, techniques for negotiating job offers, and exploration of career paths within computing fields such as industry roles, academia, research labs, or startups. The computing departments may also need to increase collaboration with the career services.

(2) Integration of soft skills and real-world problem solving. The RQ1 results show that interns' discussions covers a broad range of topics and the discussions are not limited to technical discussions. Many discussions involves social interactions and this highlights the importance of soft-skills like communication, teamwork, and problem-solving. Hence, it will be beneficial to incorporate real-world and problem-solving activities such as project-based learning into the curriculum [18]. These activities could involve team-based collaboration to simulate real-world environments, assignments requiring students to present technical work to non-technical audiences, and guest lectures from industry professionals to provide insights into industry expectations, trends, and career paths.

(3) Enhanced industry collaboration. There is a need to include industry experience as part of the curriculum. This may be in form of mandatory internships supported by partner-ships with industries to offer guaranteed placements, collaboration with industry to develop capstone projects that mirror industrial experience, and offering academic credits for participating in hackathons or coding challenges that replicate the intensity and hands-on nature of internships.

(4) Mentorship and peer support networks. The high number of posts gathered during the data collection process shows that interns often seek advice from peers. It may be necessary for academic departments to develop mentorship programs that connects students with alumni or senior peers. These programs may include peer-to-peer sessions, alumni engagement initiatives, and informal workshops where students share internship experiences, discuss challenges and celebrate success.

We believe the implementation of these guidelines will better equip students to meet the demands of the industry. The key themes suggest that emphasizing career preparation, fostering the development of soft skills, providing hands-on experience, and encouraging mentorship and networking opportunities can effectively address the challenges interns encounter in the industry. These initiatives will enhance students' readiness and competitiveness when applying for internships and professional jobs in the computing field.

7 Conclusion

This paper analyzed discussions related to computing internship on the Reddit social media platform. We extracted internship related data from June 2009 to December 2023. We then used Bert model to identify computing related posts and also used subreddits information to filter out posts that are likely to have been misclassified. The final dataset used for the analysis contained 143,912 Reddit posts.

We used topic modeling and sentiment analysis to analyze the data. The results indicate that computing students generally express a positive sentiments towards internships and tend to discuss five main topics which are academics, professional career, entertainment, social interaction, and the (software engineering) experience during internship. These findings suggest that interns consider the internship experience to be important for their academics and professional career. Furthermore, interns often navigate other aspects of life, such as social interactions, during their internships. Notably,

[6] https://stackoverflow.com/

the (RQ1) results also indicate that many of computing interns work in software development roles.

In the future, we plan to address some of the limitations discussed in Section 5, such as a deeper analysis of the posts based on the year the post was created or the demographics of the post authors. We plan to incorporate datasets from other social media platforms such as StackOverflow and compare the findings with the results presented in this paper. This will help assess whether the results are consistent across non-Reddit platforms. We also plan to analyze responses (comments) to internship-related posts to investigate whether the feedback on intern posts generally has a positive sentiment like the posts themselves. Finally, we aim to explore potential differences in internship experiences between computing and non-computing interns.

References

[1] Abdulmohsen Algarni. 2016. Data mining in education. *International Journal of Advanced Computer Science and Applications* 7, 6 (2016), 456–461.

[2] Ghassan Alkadi. 2008. Student internships: developing real world skills for real world problem solutions. *Journal of Computing Sciences in Colleges* 23, 6 (2008), 97–103.

[3] María Lucia Barron-Estrada, Ramón Zatarain-Cabada, and Raúl Oramas Bustillos. 2019. Emotion Recognition for Education using Sentiment Analysis. *Res. Comput. Sci.* 148, 5 (2019), 71–80.

[4] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. 2020. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, Vol. 14. 830–839.

[5] Jens F Binder, Thom Baguley, Chris Crook, and Felicity Miller. 2015. The academic value of internships: Benefits across disciplines and student backgrounds. *Contemporary Educational Psychology* 41 (2015), 73–82.

[6] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3, Jan (2003), 993–1022.

[7] Aaron Simon Blicblau, Tracey Louise Nelson, and Kurosh Dini. 2016. The Role of Work Placement in Engineering Students' Academic Performance. *Asia-Pacific Journal of Cooperative Education* 17, 1 (2016), 31–43.

[8] Eric Brechner. 2003. Things they would not teach me of in college: what Microsoft developers learn later. In *Companion of the 18th annual ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications*. 134–136.

[9] Jen-Chia Chang, Hsi-Chi Hsiao, Su-Chang Chen, and Tien-Li Chen. 2016. The relationship between students' background and their off-campus internship conditions for departments of electrical engineering & computer science in technological universities. *Int. J. Concept. Manag. Soc. Sci* 4 (2016), 1–5.

[10] Vahid Garousi, Gorkem Giray, Eray Tuzun, Cagatay Catal, and Michael Felderer. 2019. Aligning software engineering education with industrial needs: A meta-analysis. *Journal of Systems and Software* 156 (2019), 65–83.

[11] Vahid Garousi, Gorkem Giray, Eray Tuzun, Cagatay Catal, and Michael Felderer. 2019. Closing the gap between software engineering education and industrial needs. *IEEE software* 37, 2 (2019), 68–77.

[12] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.

[13] Arturo Jaime, Juan J Olarte, Francisco J García-Izquierdo, and César Domínguez. 2019. The effect of internships on computer science engineering capstone projects. *IEEE Transactions on Education* 63, 1 (2019), 24–31.

[14] Amanpreet Kapoor and Christina Gardner-McCune. 2019. Understanding CS undergraduate students' professional development through the lens of internship experiences. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. 852–858.

[15] Amanpreet Kapoor and Christina Gardner-McCune. 2020. Barriers to securing industry internships in computing. In *Proceedings of the Twenty-Second Australasian Computing Education Conference*. 142–151.

[16] Amanpreet Kapoor and Christina Gardner-McCune. 2020. Exploring the participation of CS undergraduate students in industry internships. In *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*. 1103–1109.

[17] Barbara Kitchenham, O Pearl Brereton, David Budgen, Mark Turner, John Bailey, and Stephen Linkman. 2009. Systematic literature reviews in software engineering–a systematic literature review. *Information and software technology* 51, 1 (2009), 7–15.

[18] Dimitra Kokotsaki, Victoria Menzies, and Andy Wiggins. 2016. Project-based learning: A review of the literature. *Improving schools* 19, 3 (2016), 267–277.

[19] Norbert J Kubilus. 2000. Assessing a computer science curriculum based on internship performance. In *Proceedings of the fifth annual CCSC northeastern conference on The journal of computing in small colleges*. 301–307.

[20] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461* (2019).

[21] Steven Loria et al. 2018. textblob Documentation. *Release 0.15* 2, 8 (2018), 269.

[22] Daniel Maier, Annie Waldherr, Peter Miltner, Gregor Wiedemann, Andreas Niekler, Alexa Keinert, Barbara Pfetsch, Gerhard Heyer, Ueli Reber, Thomas Häussler, et al. 2021. Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In *Computational methods for communication science*. Routledge, 13–38.

[23] Cynthia J Martincic. 2009. Combining real-world internships with software development courses. *Information Systems Education Journal* 7, 33 (2009), 1–10.

[24] Alexey N Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. 2019. The anatomy of Reddit: An overview of academic research. *Dynamics on and of Complex Networks III: Machine Learning and Statistical Physics Approaches 10* (2019), 183–204.

[25] Raihan Mia, Anwar Hossin Zahid, Bipul Chandra Dev Nath, and Abu Sayed Md Latiful Hoque. 2020. A Conceptual Design of Virtual Internship System to Benchmark Software Development Skills in a Blended Learning Environment. In *2020 23rd International Conference on Computer and Information Technology (ICCIT)*. IEEE, 1–6.

[26] Mia Minnes, Sheena Ghanbari Serslev, and Omar Padilla. 2021. What do cs students value in industry internships? *ACM Transactions on Computing Education (TOCE)* 21, 1 (2021), 1–15.

[27] Pan-Wei Ng and Shihong Huang. 2013. Essence: A framework to help bridge the gap between software engineering education and industry needs. In *2013 26th International Conference on Software Engineering Education and Training (CSEE&T)*. IEEE, 304–308.

[28] Joann J Ordille. [n. d.]. Internships Enhance Student Research and Educational Experiences. https://cra.org/crn/2008/11/internships_enhance_student_research_and_educational_experiences/.

[29] Oladapo Oyebode and Rita Orji. 2019. Social media and sentiment analysis: the Nigeria presidential election 2019. In *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*. IEEE, 0140–0146.

[30] Nicholas Proferes, Naiyan Jones, Sarah Gilbert, Casey Fiesler, and Michael Zimmer. 2021. Studying reddit: A systematic overview of disciplines, approaches, methods, and ethics. *Social Media+ Society* 7, 2 (2021), 20563051211019004.

[31] Alex Radermacher and Gursimran Walia. 2013. Gaps between industry expectations and the abilities of graduates. In *Proceeding of the 44th ACM technical symposium on Computer science education*. 525–530.

[32] Alex Radermacher, Gursimran Walia, and Dean Knudson. 2014. Investigating the skill gap between graduating students and industry expectations. In *Companion Proceedings of the 36th international conference on software engineering*. 291–300.

[33] Radim Rehurek and Petr Sojka. 2011. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic* 3, 2 (2011), 2.

[34] Thanveer Shaik, Xiaohui Tao, Christopher Dann, Haoran Xie, Yan Li, and Linda Galligan. 2023. Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal* 2 (2023), 100003.

[35] HBCU Summit, NACE Learning Platform, NACE Quick Polls, NACE Briefs, and Job Market. 2017. The positive implications of internships on early career outcomes. *Nace Journal* (2017).

[36] Penny Sweetser Kyburz, Alaine King, Timothy DeWan, et al. 2020. Setting Students up to Succeed in Computing Internships. (2020).

[37] Shaheen Syed and Marco Spruit. 2017. Full-text or abstract? examining topic coherence scores using latent dirichlet allocation. In *2017 IEEE International conference on data science and advanced analytics (DSAA)*. Ieee, 165–174.

[38] Joo Tan and John Phillips. 2005. Incorporating service learning into computer science courses. *Journal of Computing Sciences in Colleges* 20, 4 (2005), 57–62.