Mining mobile application reviews to inform the design of anonymous live counseling applications

Kalee Crandall, *Dakota State University, kalee.crandall@trojans.dsu.edu*
Kodey Crandall, *Utah Valley University, kcrandall@uvu.edu*
Matt North, *Utah Valley University, mnorth@uvu.edu*
Cherie Noteboom, *Dakota State University, cherie.noteboom@dsu.edu*

Abstract

Mental health issues among American youth continue to increase. Technology integration and acceptance represent modern opportunities to provide more access to licensed professionals. School counselors are mental health experts in schools uniquely positioned to provide social and emotional support to students. However, not all students have access to a school counselor or another mental health professional. In 2018, Utah developed the SafeUT app to increase services provided by school counselors and licensed clinicians to students in the state. This free mobile app provides confidential, real-time crisis intervention through an anonymous live chat and confidential tip program. Although the number of tips continue to increase each year, little research explores the design principles influencing the user's experience of anonymous live counseling applications. This research in progress paper uses data mining techniques on app reviews from the SafeUT mobile application to explore the common themes relating to design principles. Our results may inform future research and improve the design of similar applications to increase students' access to counselors and the much-needed mental health services these professionals can provide.

Keywords: counseling services, data mining, app reviews, information systems

Introduction

In October of 2021, the American Academy of Pediatrics, the American Academy of Child and Adolescent Psychiatry, and the Children's Hospital Association declared child and adolescent mental health a national emergency stating that childhood mental health concerns continue to rise. As of 2018, suicide is the second leading cause of death for youth ages 10-24 (*AAP-AACAP-CHA Declaration of a National Emergency in Child and Adolescent Mental Health*, 2021). The COVID-19 pandemic has only escalated these concerns. The American Academy of Pediatrics, American Academy of Child and Adolescent Psychiatry, and Children's Hospital Association have called for an increase in "strategies to meet these challenges through innovation and action, using state, local and national approaches to improve the access to and quality of care across the continuum of mental health promotion, prevention, and treatment" (*AAP-AACAP-CHA Declaration of a National Emergency in Child and Adolescent Mental Health*, 2021).

Students who are physically, mentally, and socially healthy are better learners. Gretchen Whitmer, Governor of Michigan, said, "Having skilled professionals in school buildings helps our kids get the supports they need so they can thrive in the classroom and beyond "(*Michigan Schools Use New State Grant to Invest in Students' Mental and Physical Health, Recruit 560+ Nurses, Social Workers, Counselors*,

2021). School counselors provide critical social-emotional and academic support. Through a comprehensive school counseling program that promotes success and achievement for all students, school counselors can also help set students on a path for post-secondary education and help them achieve success in their future careers. Despite these services, nearly one in five students (roughly eight million children) in the U.S don't have access to a counselor in their school at all, and nearly three million of those students don't have access to other school support staff, such as a school psychologist or social worker (Education Trust, 2019).

In 2018, the Utah State Legislature and the Utah Office of the Attorney General funded the development of the SafeUT app to increase services provided by school counselors and licensed clinicians to students across the state. This free mobile app provides students with confidential, anonymous, real-time crisis intervention through a live chat with a licensed counselor. Although the number of users of this app continue to increase each year, questions remain about whether or not it is improving access to mental health specialists and meeting the needs of its users.

Live chat systems have many benefits. Research has shown that these systems can lower interaction costs, provide critical support, foster positive relationships (Krasnova et al., 2017), increase customer satisfaction due to 24-hour support, increase levels of trust, and encourage repeat visits (McLean & Osei-Frimpong, 2017). Despite all of the benefits of these systems, limited research examines the variables influencing satisfaction with the live chat experience (McLean & Osei-Frimpong, 2017). To date, the authors found that no study has explored these principles in an anonymous live chat like that found on the SafeUT app.

Mobile application reviews offer real-time data about users' opinions of the systems. These reviews are easily accessible and enable researchers to extract users' preferences instantly, leading to better design and development of similar systems.

The purpose of this exploratory study is to use the k-means clustering technique to analyze mobile application reviews of the SafeUT application and answer the research question: *"What are the design principles influencing user satisfaction of anonymous mobile chat systems?"* Answering this question may lead to a greater understanding of design principles influencing user acceptance and usability of anonymous live counseling applications to increase access and services provided by counselors.

Related Work

Previous research has focused on the benefits of analyzing users' reviews to inform design principles (Al-Ramahi et al., 2017; Ciurumelea et al., 2017) the usefulness of the k-means clustering technique to extract text features (Debao et al., 2021; Wu et al., 2008), and variables that influence user satisfaction in live chat services (McLean & Osei-Frimpong, 2017).

Analyzing users' reviews of mobile applications is a newer method for gathering feedback from a system's user. This technique informs developers of design principles influencing consumers' preferences and provides knowledge to enhance their efficacy (Al-Ramahi et al., 2017). However, while these reviews contain valuable feedback directly from the app users, "the amount of unstructured text information they include can be overwhelming for developers" (Ciurumelea et al., 2017, p. 1).

The k-means clustering algorithm is an unsupervised learning technique that partitions large amounts of data, including data derived from text documents, into groups (Abualigah & Khader, 2017). Clustering methods find similarities between points of data and function by portioning large datasets into different

small subgroups or clusters based on a similarity measure (Kulis & Jordan, 2012). Researchers Debao, Yinxia, and Min (2021) used a semantically based k-means clustering algorithm to obtain recruitment information related to big data on Zhaopin.com, a recruitment website. Their k-means clustering analysis effectively demonstrated wage differences among different jobs and the importance of education and experience in wage settings. These authors state that "k-means clustering can effectively understand the passage, process natural language from all aspects, and dig out its potential information and value" (Debao et al., 2021, p. 2).

Live chat services are a cost-effective way to provide 24-hour customer assistance (McLean & Osei-Frimpong, 2017). These systems allow individuals access to a service representative online, from any device, at any time to ask a question and receive a response. McLean and Osei-Frimpong (2017) examined the satisfaction of a user's experience during a live chat service using structural equation modeling. The authors found that the most influential constructs of these systems were service quality, information quality, and system quality, all of which were dependent on the purpose of use. This study also found that the "role of emoticons, presence of service reps picture, automated 'canned' responses and the presence of response time estimations" had only a moderate influence on the constructs and encouraged future research to test their model (McLean & Osei-Frimpong, 2017, p. 1).

Research Methodology

To explore the design principles influencing user satisfaction of anonymous mobile chat systems, this research uses a k-means text clustering approach to identify design themes. K-means clustering is a process of unsupervised machine learning that uses a clustering algorithm to learn from data properties for labeling groups of data points. While there are many approaches and systems that allow for in-depth k-means analysis, this study uses an open-source algorithm available from the k-means library in Python.

Data Collection and Preparation

The first step in analyzing mobile application reviews is data collection. SafeUT application reviews from iTunes (Apple) and the GooglePlay store were exported from AppFollow.io. This online web application is a flexible platform that allows users to "follow" application reviews as they are submitted to top app stores. Data exported from AppFollo.io for this research included 105 SafeUT app reviews from Apple iTunes and 211 SafeUT app reviews from the GooglePlay store.

Once the data was collected, it was stored in Comma Separated Values (.csv) form so that the app reviews could be imported into Jupyter Notebook (J.N.). Jupyter Notebook is an open-source web or desktop application used to create and share documents that contain live code, equations, visualizations, and text. We used J.N. to compute the k-means clustering of the app reviews. We also used J.N. to identify any invalid or incomplete data. We found that no reviews were missing any data. Although some reviews had very few words, we decided to use them in the analysis as app reviews typically do not contain an extensive amount of qualitative data (Dhar & Bose, 2022). Additionally, J.N. allowed us to prepare the data by automatically filtering out auxiliary and meaningless words such as "and, the, it," removing symbols and any other irrelevant characters.

The textblob library in Python was then used to perform sentiment analysis. This allowed us to compare the text reviews to determine polarity and subjectivity. The polarity exhibits positive or negative responses based on the words used in a review. Subjectivity demonstrates the intensity of the emotion in the review.

Step 3: K-Means Cluster Analysis

K-means text clustering is commonly used for text analysis (Debao et al., 2021) as the algorithm is simple, easy to understand, and reasonably scalable (Wu et al., 2008). The k-means library in Jupyter Notebook was used to run the analysis and to organize the dataset into 9 clusters.

The optimal number of clusters was identified using the Elbow Method Algorithm (EMA). The EMA squares the distance between points in each cluster and the cluster's centroid to give a series of K values. The sum of squared errors (SSE) is then used as a performance indicator and helps identify the optimal number of clusters (Yuan & Yang, 2019).

The clusters were reviewed by the first and second authors, who then performed some posthoc analysis of the text mining results. Previous research using text mining demonstrates that sometimes a manual process is required to make sense of the analyzed word tokens (Plaisant et al., 2006).

After the clusters were thoroughly reviewed, keywords were then placed into themes. A label was given to each theme based on a topic that relates to specific design principles.

Table 1: Design Principle Linked to Cluster Analysis

Usefulness	Intention to Use	Features	Ease of Use	Service Quality
Great	Required	Like	Love	Helpful
App	Thanks	Just	App	Helps
Helped	App	Really	Really	Problems
Thank	Asks	Know	School	Depressed
Amazing	Use	Chat	Schools	Need
Friends	Told	App	Recommend	A lot
Tell	Choose	Time	Actually	Haven't
Works	Wanting	Make	Helpful	Thing
Counselor	Attend	Text	Find	Time
	Forced	Call	Easy	
		Option		

Service Provided	Information Quality	Connection	System Quality
Talk	App	Good	Crisis
Feel	School	Idea	Download
Bad	Use	Help	App
Try	Helped	Platform	Safe
Things	Amazing	Shy	Line
Chat	Life	Feel	Boring
Didn't	School	App	List
Conversation	Best	Struggling	Bugs
Worked		Compassion	Glitching
			Freezes

Step 4: Distribution Analysis

In addition to mining for keywords in the app reviews, we evaluated the star ratings given by each reviewer. There were 62 one-star ratings, 15 two-star ratings, 18 three-star ratings, 23 four-star ratings, and 93 five-star ratings, as shown in Table 2 and Figure 1.

Table 2: Frequency of Star Ratings

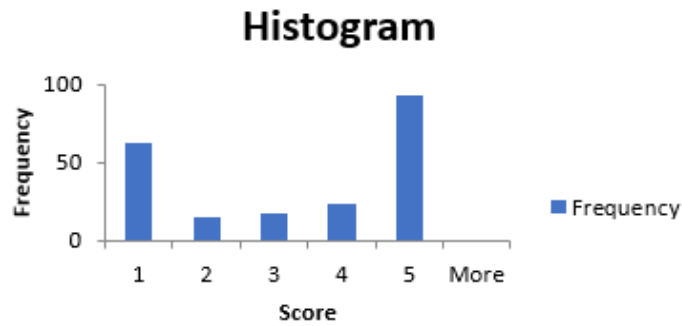| Score	Frequency
1	62
2	15
3	18
4	23
5	93

Figure 1: Histogram of Star Ratings

Step 5: Sentiment Analysis

The sentiment analysis shows extremes for positive and negative reviews, which can be expected for those with stronger emotions (Langhe et al., 2016). This analysis was performed in J.N. using the textblob library in Python.
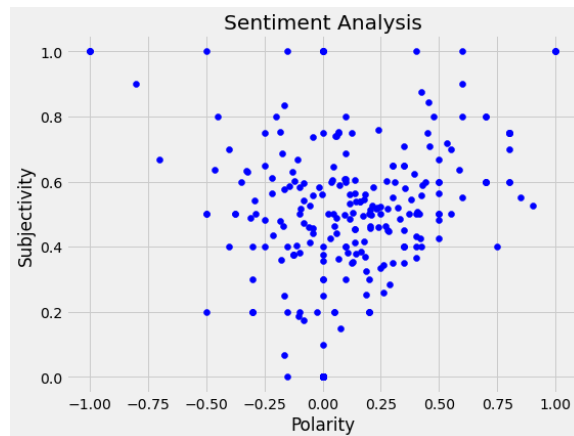

Figure 2: Subjectivity and Polarity

Findings

Research on the validity and power of product rating scales (often one to five stars, as found in our data) shows that these ratings are often biased but influential for consumer technology acceptance (Langhe et al.,

2016). The bi-modal distribution typical in 1–5 star rating systems is present in this dataset as seen in the frequency of star ratings provided by the users (Table 2) and the polarity of the reviews as calculated by J.N. in the sentiment analysis (Figure 2).

Some words appear in multiple categories of the k-mean cluster analysis. This is common in textual cluster analysis using individual words. The authors looked at the compilation of each term in a cluster to create one central theme relating to a design principle. The themes are Usefulness, Ease of Use, Intention to Use, Features, Service Quality, Service Provided, Information Quality, Connection, and System Quality.

Usefulness

Usefulness is one of the key constructs of the Technology Acceptance Model (TAM) proposed by Davis (1989). In this theory, behavioral intention (B.I.) to use a system is determined partly by a product's perceived usefulness, which is the extent to which a user believes that the system will enhance his or her job performance. For SafeUT, perceived usefulness refers to the user's belief that the app would get them the support needed. Examples in the dataset include the following:

> *"I wanted to kill myself but this helped me very much thank you"*

> *"This app helped me so much when i was having panic attacks, mental issues, and depression"*

Ease of Use

Perceived ease of use is another construct of the TAM (Davis, 1989). It is defined as "the degree to which a person believes that using a particular system would be free from effort" (Davis, 1989). Examples in the dataset include the following:

> *"This app, easy access to it, and people who encourage you to use the app -- has saved my life many times, and MANY other lives of the people i know. I'm a 17 yr old young woman in a high school in Utah, and I've personally attempted 3+ times as of two years ago, and so many of my peers have too. Use this app please."*

> *"I 100% recommend. Real people (like professionals) reply back to your messages. It's a good outlet to have if you have anxiety or ever need someone to talk to. It's a good source to just vent overall. No matter what it is that you're talking about. they offer help services for more serious (or what they feel are serious) "sessions"."*

Intention to Use

Intention to use is defined as "the degree to which a person has formulated conscious plans to perform or not perform some specified future behavior" *(Brezavšček et al., 2016)*. Some examples of this in the dataset include:

> *"You press a few buttons and get emotional support. That's something I've wanted for a long, long time. Don't take something like this for granted; this is the kind of thing made to help you in the toughest situations."*

> *"The school forced us to download it."*

Features

Mobile application features are user interface (U.I.) enhancements that allow the user to interact with and access the content (Chen et al., 2019). Examples of this in the dataset include:

> *"I really needed some advice and this was perfect! The app allows you to text/and call professionals who know what to do. They're sincere to whatever your situation or question is and they actually help! Everything is anonymous, free of charge, and it doesn't ask for any personal information. I love it!!!"*

> *"How come there's no "teacher" status option, only student and parent?"*

> *"As a parent, I would like the option to set up an alert or notification for the schools my children attend if a tip or chat were occurring so that I could know what is going on."*

Service Quality

Service quality is defined as "the differences between expectations and performance along the dimensions of quality" (Parasuraman et al., 1988). This design principle relates to the expectations of the users of SafeUT and the quality of the app's performance. Some examples of this in the dataset include:

> *"Has great potential, but it takes a really long time for anyone to reply to you. If they could make it so they answered you sooner and faster, you could make sure that more people were saves from anything from having a bad day to having issues with suicidal thoughts."*

> *"I love it. Very helpful"*

Service Provided

The design principle, "Service Provided" relates to the actual service provided by the SafeUT app. Examples of this in the dataset include:

> *"I reported my friend and she got help with in a few days it worked and I am happy with it."*

Information Quality

Information quality refers to the quality of the information included in the SafeUT app. A system is said to have high information quality when the information being presented is consistent in meeting the user's expectations (Office of the Chief Information Officer, 2003). An example of this in the dataset is:

> *"My sons school is not in the list of schools."*

Connection

Connection is an important design principle critical for the success of an app like SafeUT. Many individuals who use SafeUT are considering self-harm. When a system incorporates the design principle of "Connection," users are more likely to "change their behavior and achieve their goals as they are more likely to perform when they perceive social support and observe others' performance" (Al-Ramahi et al., 2017, p. 19). Examples of this in the dataset include:

> *"What a great/supportive group of caring therapists. So much better than the national crisis line. These folks are compassionate and this app is a godsend. UNI is the best!"*

> *"Good for people who are shy"*

System Quality

System quality refers to the perceptions of the app's users. The app may be unsuccessful if the user finds the system helpful but does not find it aesthetically appealing. Additionally, if the app looks good but does not function as expected, system quality is compromised, and the mobile application will not be accepted (Kim et al., 2015). Examples of this in the dataset include:

> *"This app is boring"*

> *"The concept, and outline of the app is great, but the implementation is poor. The developers did not put enough time in. Lots of bugs, for example: If the design was more functional (and more professional visually) it would be great. I am also still warry on allowing anonymous reporting because it has high potential for abuse."*

> *"I can't download this right now, my app is glitching, but i need help…"*

> *"App freezes and it goes to Parent or guardian which AMA client."*

The frequency of the distributions reveals just under 75% of the ratings fall into the best or worst categories.

Discussion

Our k-means analysis of these app reviews revealed both positive and negative elements of technology acceptance for the SafeUT app. Design principles of Usefulness, Intention to Use, Ease of Use, Information Quality, and Connection all contained predominantly positive keywords in the reviews we mined. Post-hoc analysis of comments in these design areas confirmed these findings: reviewers primarily praised the SafeUT app's design in these areas. However, our clustering analysis also revealed some negativity toward design principles of Service Quality, Service Provided, and System Quality, with mixed comments on Features and Users. Thus, there is room for improvement in SafeUT's design to improve acceptance of the app and, more importantly, access to the services available to students through its use.

This division of acceptance across the nine design principles maps to the users' star ratings. About 50% of the design principles were associated with positive keywords in the text analysis, and just less than 50% of the star ratings given were 5/5 stars. About 30% of design principles were associated with negative words, and 30% of the star ratings were 1/5 stars. There were a few 'middling' design principles based on keywords found, along with a little more than 20% of the reviews receiving middling star ratings of 2, 3, or 4/5 stars.

Though both the written reviews and star ratings are somewhat subjective, our analysis does give insight into the reviewers' perceptions of SafeUT's strengths, weaknesses, and opportunities for improvement. We specifically find that developers of anonymous live counseling systems would do well to focus on improving Service and System Quality.

Future Research

The need for access to counselors has never been greater. The number of counselors is limited; however, access to counselors through anonymous, online technology is enhanced through mobile applications like SafeUT. Future research can enhance our findings by examining specific design principles for effectiveness and acceptance by end-users. The k-means cluster approach was initially helpful in identifying themes and mapping these to design principles. Future research should use another statistical approach to validate the results and identify other design features to increase technology acceptance.

Conclusion

The increasing prevalence of mental health concerns in youth, coupled with the limited access to innovative solutions to address these challenges calls for innovative solutions to address these challenges. The SafeUT application aims to address these challenges, by providing students with confidential and anonymous real-time crisis intervention through live chat with licensed counselors. However, there is a need to understand the design principles that influence user satisfaction with such anonymous mobile chat systems. Using the k-means clustering technique, we identify several design principles influencing the user satisfaction of the SafeUT app: Usefulness, Ease of Use, Intention to Use, Features, Service Quality, Service Provided, Information Quality, Connection, and System Quality.

The perceived usefulness of the app was highlighted by users who expressed gratitude for the support they received during challenging times. The ease of use was another significant factor, with users appreciating the simplicity and accessibility of the app. Additionally, the intention to use the app was influenced by the presence of real professionals who replied to messages, providing a valuable outlet for venting and seeking advice. Features such as text and call options, as well as the anonymity and confidentiality of the app, were also positively mentioned by users.

The study underscores the importance of these design principles in enhancing user satisfaction with anonymous mobile chat systems. By understanding these factors, developers can further improve the efficacy and usability of such applications. The analysis of mobile application reviews proves to be a valuable method for gathering real-time feedback and informing the design and development of similar systems.

Overall, this research contributes to the broader goal of increasing access to mental health support for children and adolescents. By leveraging technology and understanding user preferences, it is possible to provide effective and user-friendly solutions that address the mental health needs of young individuals, ultimately fostering their overall well-being and academic success.

References

AAP-AACAP-CHA Declaration of a National Emergency in Child and Adolescent Mental Health. (2021, October 19). http://www.aap.org/en/advocacy/child-and-adolescent-healthy-mental-development/aap-aacap-cha-declaration-of-a-national-emergency-in-child-and-adolescent-mental-health/

Abualigah, L., & Khader, A. (2017). "Unsupervised text feature selection technique based on hybrid particle swarm optimization algorithm with genetic operators for the text clustering." *Journal of Supercomputing*, *73*(11), 4773–4795. https://doi.org/10.1007/s11227-017-2046-2

Al-Ramahi, M. A., Liu, J., & El-Gayar, O. F. (2017). "Discovering Design Principles for Health Behavioral Change Support Systems: A Text Mining Approach." *ACM Transactions on Management Information Systems*, *8*(2–3), 5:1-5:24. https://doi.org/10.1145/3055534

Brezavšček, A., Šparl, P., & Žnidaršič, A. (2016). "Factors Influencing the Behavioural Intention to Use Statistical Software: The Perspective of the Slovenian Students of Social Sciences." *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(3). https://doi.org/10.12973/eurasia.2017.00652a

Chen, X., Zou, Q., Fan, B., Zheng, Z., & Luo, X. (2019). "Recommending software features for mobile applications based on user interface comparison." *Requirements Engineering*, *24*(4), 545–559. https://doi.org/10.1007/s00766-018-0303-4

Ciurumelea, A., Schaufelbuhl, A., Panichella, S., & Gall, H. C. (2017). "Analyzing reviews and code of mobile apps for better release planning." *2017 IEEE 24th International Conference on Software Analysis, Evolution and Reengineering (SANER)*, 91–102. https://doi.org/10.1109/SANER.2017.7884612

Davis, F. D. (1989). "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS Quarterly*, *13*(3), 319. https://doi.org/10.2307/249008

Debao, D., Yinxia, M., & Min, Z. (2021). "Analysis of big data job requirements based on K-means text clustering in China." *PLOS ONE*, *16*(8), e0255419. https://doi.org/10.1371/journal.pone.0255419

Dhar, S., & Bose, I. (2022). "Walking on air or hopping mad? Understanding the impact of emotions, sentiments and reactions on ratings in online customer reviews of mobile apps." *Decision Support Systems*, 113769. https://doi.org/10.1016/j.dss.2022.113769

Education Trust. (2019). *School Counselors Matter*. Education Trust https://www.schoolcounselor.org/getmedia/b079d17d-6265-4166-a120-3b1f56077649/School-Counselors-Matter.pdf

Kim, D., Park, J., Kum, S. W., & Chung, K. (2015). A context-aware matching system to improve user-perceived quality *IEEE Transactions on Consumer Electronics*, *61*(4), 531–538. https://doi.org/10.1109/TCE.2015.7389809

Krasnova, H., Veltri, N. F., Eling, N., & Buxmann, P. (2017). Why men and women continue to use social networking sites: The role of gender differences *The Journal of Strategic Information Systems*, *26*(4), 261–284. https://doi.org/10.1016/j.jsis.2017.01.004

Kulis, B., & Jordan, M. I. (2012). Revisiting k-means: New algorithms via Bayesian nonparametrics *Proceedings of the 29th International Coference on International Conference on Machine Learning*, 1131–1138.

Langhe, B. de, Fernbach, P., & Lichtenstein, D. R. (2016, July 4). High Online User Ratings Don't Actually Mean You're Getting a Quality Product *Harvard Business Review*. https://hbr.org/2016/07/high-online-user-ratings-dont-actually-mean-youre-getting-a-quality-product

McLean, G., & Osei-Frimpong, K. (2017). Examining satisfaction with the experience during a live chat service encounter-implications for website providers *Computers in Human Behavior*, *76*, 494–508. https://doi.org/10.1016/j.chb.2017.08.005

Michigan Schools Use New State Grant to Invest in Students' Mental and Physical Health, Recruit 560+ Nurses, Social Workers, Counselors. (2021, December 28). https://www.michigan.gov/whitmer/0,9309,7-387-90499_90640-574659--,00.html

Office of the Chief Information Officer. (2003). Total Information Quality Management Handbook *U.S. Department of Housing*, *3300.1*.

Parasuraman, A., Zeithaml, V., & Berry, L. (1988). Servqual: A Multiple-Item Scale For Measuring Consumer Perc *Journal of Retailing*, 12.

Plaisant, C., Rose, J., Yu, B., Auvil, L., Kirschenbaum, M. G., Smith, M. N., Clement, T., & Lord, G. (2006). Exploring erotics in Emily Dickinson's correspondence with text mining and visual interfaces. *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*, 141–150. https://doi.org/10.1145/1141753.1141781

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1–37. http://dx.doi.org/10.1007/s10115-007-0114-2

Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. *J*, *2*(2), Article 2. https://doi.org/10.3390/j2020016