June 2024

An AI Approach for Predicting Audience Reach of Presentation Slides

Alexander Meier
University of St. Gallen, alexander.meier@unisg.ch

Roman Rietsche
Bern University of Applied Sciences, roman.rietsche@bfh.ch

Ivo Blohm
University of St. Gallen, ivo.blohm@unisg.ch

AN AI APPROACH FOR PREDICTING AUDIENCE REACH OF PRESENTATION SLIDES

Short Paper

Alexander Meier, University of St. Gallen, St Gallen, Switzerland, alexander.meier@unisg.ch
Roman Rietsche, Bern University of Applied Sciences, Bern, Switzerland, roman.rietsche@bfh.ch
Ivo Blohm, University of St. Gallen, St Gallen, Switzerland, ivo.blohm@unisg.ch

Abstract

There is a near overflow of presentation slides on digital platforms, such as SlideShare.net, with 40 million. This presents a challenge in assessing their projected impact due to its high complexity and required expertise. We propose a novel approach using machine learning techniques to predict presentation slide audience reach. We crawled a unique dataset of over 8000 slides and extracted relevant attributes. A model was trained where we are the first to employ both numerical and textual inputs. Initial results with an R^2 value of 0.579 suggest that the audience reach of presentation slides can be automatically evaluated. Our findings contribute to the current understanding of the assessment of online documents, introducing possibilities for further research, such as focusing on domain-specific applications and incorporating them as tools for decision support in content management systems on sharing platforms.

Keywords: Machine Learning, Natural Language Processing, Slide Evaluation, Decision Support.

1 Introduction

According to a 2016 poll, an estimated 500 million users deliver around 35 million PowerPoint presentations daily (polleverywhere, 2016). Platforms such as SpeakerDeck.com and SlideShare.net have established repositories catering to varied audiences and thematic scopes. For instance, SlideShare.net boasts a repository of 40 million presentations, attracting 100 million users monthly (Ha, 2020). Plenty of platforms offer this kind of shared repositories, however with varying success in attracting visitors. Latest developments in Large Language Models (LLM) lowered access barriers to ready-to-use presentations even further. The professional, academic, and educational audiences deploy these platforms to gain fresh perspectives, incorporate diverse approaches, and potentially enhance the impact and relevance of their presentations (Thelwall and Kousha, 2016).

There are several existing approaches aiming at automated slide or even presentation generation. Sefid et al., 2019 employ deep neural networks for slide creation using extractive summarization, though this might limit capturing complex content nuances. Cagliero and Quatra, 2021 present an unsupervised method for scenarios with limited training data, offering flexibility but struggling with complex materials. Fu et al., 2022 in contrast tackle multimodal summarization and structured output, standing out for its comprehensive approach, yet facing scalability challenges due to its complexity.

Existing literature highlights the need for a balanced model that efficiently handles complexity, scalability, data availability and serves as a valuable decision support system (DSS), aiding in evaluating presentation slides. This includes platforms desiring to filter and eventually offer high impact slide decks. Such a solution is greatly needed as the performance of LLMs to generate high impact presentations is still not reliable (Guo et al., 2023), particularly considering the risk of these AI-generated presentations flooding platforms, which could lead to a dilution of content attractivity. Secondly, amidst the large number of visitors and resources available, it is challenging to follow the

determinants behind some successful slide decks that capture attention, while others remain entirely unnoticed. The evaluation of presentations is manifold and requires a high degree of didactic and subject specific expertise. For users, the complexity of choosing, navigating, and selecting themselves is high due to the huge amounts of data. Concurrently, the prospective impact and scope of sharing lecture slides—enabling broader dissemination of knowledge and fostering engagement—remain ambiguous before uploading. One primary goal is to empower content creators and platforms to reuse, create, or identify presentations that attract high possible audience reach and impact. Hence, a computational approach emerges as a solution capable of delivering insightful results while maintaining scalability. Therefore, this paper investigates the following research questions (RQ):

RQ1: *What features are decisive for the measuring audience reach of presentation slides?*

RQ2: *To what extent can the audience reach of presentation slides be assessed automatically?*

Machine learning (ML), a subset of artificial intelligence (AI), automates data analysis and forecasting and has recently advanced significantly. ML techniques enable automated, scalable assessment and rating content by quickly extracting insights from visual and textual elements (Rhyn and Blohm, 2017). Talks and presentations offer fertile ground in multiple domains for the application of ML methodologies. Such platforms benefit from abundant training sets, commonly available from slide libraries, and can leverage pre-existing models for new contexts, thus making them versatile across different thematic settings.

To answer the RQs we (1) crawled a large dataset of over 8,000 slide decks, (2) selected and engineered inductive features based on qualitative analysis of the slides and mapped them onto corresponding categories and dimensions deductively based on existing literature, and (3) trained and evaluated a meta-model. Moving forward, our research will focus on two main areas. First, we plan to augment the meta model by incorporating a broader range of features and adapting it for varied domains like corporate presentations and educational lectures. Second, we aim to empirically validate the model by integrating it into content management platforms for real-time analysis and conducting experimental studies to test its utility in reducing the cognitive load of evaluating presentations.

Addressing RQ1, we contribute with a new meta model in identifying engaging presentations. The results showed the ability to regress the user engagement of presentation slides with an R2 score of 0.579 using natural language processing (NLP) capabilities to augment yet solely feature-engineered models.

2 Background and Related Work

In this, features in digital slide decks and current methods for assessing their corresponding impact are discussed. We further briefly outline the strengths and shortcomings of existing automated approaches.

2.1 Features in presentation slides

Quality in presentation slides, whether for talks, presentations, or education, has been widely researched. Drapkina et al., 2022 highlight the importance of basing content on pedagogical principles and user priorities. According to them content, structure, and logic of the material presentation are significant factors affecting satisfaction with participation in educational events. Hosseini, 2022 outlines the slide's content should be based on valid, trustworthy resources and be adapted to the audience's needs and characteristics. In the following the understanding of Information Quality (IQ) focused on presentation slides should be guided by the taxonomy of Kim et al., 2017. They differentiated between intrinsic, representational, contextual, and reputational categories. Intrinsic content is determined by its accuracy and cohesiveness, which are paramount, to ensuring error-free and relevant content. Representational focuses on clear and consistent presentation of information. The Contextual aspect covers the completeness and current relevance of the material. Lastly, the Reputational category considers the credibility conferred by the author or institution's standing. On online platforms like slideshare.net or speakerdeck.com, it is assumed that these categories directly translate into the number of clicks, or better the number of views of a shared presentation.

2.2 Automated slide evaluation approaches

In literature, few scholars have used automated approaches to assess digital presentation slides. Luzardo et al., 2014 used machine learning models to analyze 448 presentations, classifying students as high or low performers based on features such as the number of words, images, tables, and audio aspects such as pitch. They achieved an accuracy of up to 65% for slides and 69% for audio features. They advocate for using simple features for evaluating presentation skills. Yi et al., 2022 focused on single slides, employing a neural network to differentiate between novice and well-designed slides using visual and structural features. They worked with 1,080 slide pairs, each consisting of an original novice slide and its improved version based on expert advice. By addressing the class imbalance in design problem datasets and applying resampling methods and multitask learning, they achieved an average accuracy of 81.79% and added additional insights to their system in enhancing novice slide design skills.

Therefore, research has shown that some first automated classifiers perform slide deck evaluation tasks to a satisfying degree. To the author's knowledge, all previous approaches to such computational evaluations were based on features designed by humans. These features can be a potential bottleneck, requiring significant domain expertise to design (Deng and Liu, 2018). An NLP approach can holistically analyze textual content to accurately identify patterns indicative of a high audience reach slide deck. However, this simplifies a complex issue, where using a high dimensionality of features may lead to overfitting, causing the model to learn training set-specific noise rather than generalizable features for real-world applications (Zhang et al., 2017).

We, in contrast to prior research, settled on two distinct kinds of inputs. These are (1) what we call numeric features, which have been to some extent used in previous approaches like the number of pictures, number of words, number of pages, etc., and (2) the entire textual content, which to the author's best knowledge has not been used in any prior approach.

3 Research Methodology: Model Development and Generation of Experimental Training Data

In this study, we design, train, and evaluate multiple DL and ML models with different inputs to investigate if they can be used for automated slide evaluation. Following the guidelines for conducting ML research in IS (Kühl et al., 2021), we give a detailed account of our data analysis in the method and results section.

3.1 Data crawling and variables

The data for this study was crawled from the slidesharing platform “speakerdeck.com”. The crawler was built to harvest a significant number of slide decks with maximum diversity in terms of topics, languages, and dedicated audiences. In a random approach, the crawler retrieved 7,945 URLs from the platform after removing duplicates. Subsequently, the slide decks were retrieved in the PDF format together with the corresponding metadata, which includes the title, a unique identifier for the author, a topic category, and the number of views that will serve as our dependent variable. The number of views objectively quantifies user engagement, reach, and impact of content. The measure is suited for benchmarking, allowing comparisons that inform content optimization. Importantly, they offer stakeholders a clear, even monetizable metric.

By parsing the PDFs, using the open-source "pypdf” library, we were able to extract further data variables together with the crawled metadata (Fenniak et al., 2022). The comprehensive data collection process ensured that multiple categories and corresponding dimensions are mapped to be assessed by multiple tailored features. The corresponding mappings of the categories to the variables are displayed in Table 1.

Variable	Description	Category (Dimension)	Mean	Std	Max	Min
content	The textual content of all	Intrinsic	NA (String)			

	slides in a presentation	(Accuracy, Cohesiveness)				
category	The content category the presentation belongs to	Contextual (Relevance)	NA (String)			
views	The total number of views and visits to the slide deck	NA/ Dependent Variable	4827.6	32921.4	905239	0
mean_numWords	The mean number of words per slide	Representational (Clarity) and Contextual (Informativeness)	28.6	47.1	1464.2	1.0
std_numWords	The standard deviation of the number of words across all pages in the slide deck		26.1	162.7	9757.1	1.0
total_numWords	The total number of words in the slide deck		1241.8	4247.8	218945	1
entropy	Entropy calculated from word frequencies, scaled by page count and max frequency	Intrinsic (Cohesiveness)	255.3	708.7	34324.4	1.3
numPages	The number of pages in the slide deck	Contextual (Informativeness)	50.8	55.4	1181	1
Page Density Index (PDI)	$1 - \frac{numPages}{total\_numWords}$		-26.1	42.2	1.0	-1420.1
total_numImages	The total number of images in a slide deck		174.3	2027.8	143674	0
std_numImages	The standard deviation of the number of images in a slide deck		4.8	77.8	6080.7	0

Table 1. Variables used for slide deck prediction and descriptive analysis.

3.2 Data preparation

For deploying the data in training the model, preparatory steps were conducted. For this reason, cleaning steps were implemented. While the proprietary pptx file format is well a structured compressed folder using a standardized XML format that defines shapes, text, colors, and much more, PDFs do not follow a defined structure. Sometimes text is not saved as text but as an image or even the whole slide's content is saved as an image which makes parsing, extracting, and processing a lot more difficult. The unstructuredness of PDFs is a well-recognized problem in data extraction and analysis (Tekin et al., 2022). Therefore, the cleaning steps are mainly outlined by exclusion criteria, which are defined as follows:

As working with real-world data, some PDFs were corrupted. PDFs that could not be opened were neglected. As discussed later, the NLP approach for analyzing plain text requires calculating the embeddings. The model used for calculating embeddings has a token limit of 8,192, which serves as the upper limit for textual content in presentations. It is, of course, possible to navigate beyond the token limit. However, as noted in contributions on transformer models (Vaswani et al., 2023), the attention mechanism which is a core component of these models, is designed to operate within a certain scope of context. Pushing beyond this scope can strain the model's ability to maintain coherence. We further excluded presentations with a total picture-to-word count ratio of more than two. A higher ratio is a strong indicator that the main textual content is saved as a picture in the PDF. Further filters are a minimum of 20 entries per category or author, and the number of views greater than zero. These filters turned out to be crucial in mitigating class imbalance, a known issue in machine learning for

obtaining useful predictions. Through this rigorous cleaning process, our final dataset was narrowed down to 3,459 presentation sets. The reduction in dataset size was a result of eliminating presentations that did not meet the aforementioned criteria, ensuring both the quality and the usability of our dataset for training a robust ML model.

The dependent variable, the number of views, exhibited a skewed distribution, typical of exponentially distributed variables. To address this and enhance the predictive accuracy of our machine learning model, we applied a log transformation, as is common in literature (Siddiqi and Pak, 2021). This transformation effectively normalized the distribution, reducing skewness and making the data more amenable to the assumptions of our ML and DL models.

3.3 Model training

Our modeling approach consisted of two stages. Initially, we selected ML and DL models in addition to their corresponding best-performing parameters. Subsequently, we trained the models to regress the level of attention in terms of views received. Two distinct modeling approaches were deployed, trained, and evaluated for their respective purposes. The first model leverages textual content as its input while the other model deploys manually engineered, as we call them, numerical features. These numerical features augment the regression capabilities in providing additional informational value that would not be present in sole textual description such as the number of pictures or the category the presentation belongs to. In the crawled metadata the number of stars, that users generate while using the platform, is intently not included as it could serve as a complementary proxy for impact while exhibiting a high correlation with the number of views.

The numerical data was fed into a Support Vector Machine Regressor (SVR). The ML algorithm operates by identifying the optimal hyperplane in the feature space. This hyperplane is optimized to be chosen to minimize the generalization error, leveraging the structural risk minimization principle. SVR achieves its objective by fitting a model within a specified tolerance margin around the regression line, using the support vectors to define this margin and kernel functions to manage non-linear relationships in the data (Awad and Khanna, 2015). During training, hyperparameter optimization was executed for the SVR, leading to the selection of a linear kernel. The epsilon value of 0.5 defined a moderate margin of tolerance, balancing sensitivity to data variations without overfitting. The regularization parameter, set at 200, ensured a strict fit to the training data, prioritizing minimum error over model simplicity.

In contrast, the textual content was injected into OpenAI’s proprietary GPT-3.5 called the “text-embedding-ada-002” model for generating embeddings. Text embeddings transform intricate, high-dimensional textual data into more manageable, lower-dimensional forms, effectively retaining key relationships and structural elements (Liu et al., 2020). Depending on the input and task “text-embedding-ada-002” achieves state-of-the-art performance (Balikas, 2023). The generated vector still represented a high dimensional feature space due to the size of the training set which can be an issue for the performance of ML algorithms. In this line, we applied a learning technique for dimension reduction called UMAP which is not limited in the number of features and exhibits remarkable runtime performance (McInnes et al., 2020). With UMAP the final feature space was reduced to 100 as input into a Random Forest Regressor (RFR). The RFR was fine-tuned through hyperparameter optimization, resulting in a configuration that significantly enhanced the model's performance. We chose 200 trees in the forest, providing a robust set of predictions while balancing the computational load. We specified 10 as the minimum number of samples required at a leaf node, which helped prevent overfitting by ensuring that each leaf had a reasonable number of instances. We limited the maximum depth of the trees to 10, which ensures control overfitting while allowing the model to learn sufficiently from the data.

3.4 Model architecture

As depicted in Figure 1, the architecture comprises three models in total. The first two models have already been described in the previous section. The third represents a so-called stacking "ensemble," integrating and weighting the strengths and weaknesses of the two individual learners to obtain a

superior regressor that outperforms them all (Kotsiantis et al., 2006). The output of the two baseline regressors is passed to a Multi-Layer Perceptron (MLP). The only hidden layer was set to have five neurons suitable for capturing underlying patterns while avoiding overfitting. The L-BFGS solver was chosen for its efficiency for a smaller set of features and fast convergence (Nocedal and Wright, 2006).

To build an unbiased ensemble regressor, the initial dataset was split into three parts, train, validation, and test. The two baseline regressors used the training set for training and the validation dataset for evaluation. The predictions on the validation data are then deployed to train the ensemble learner. The final evaluation is conducted on the test dataset.

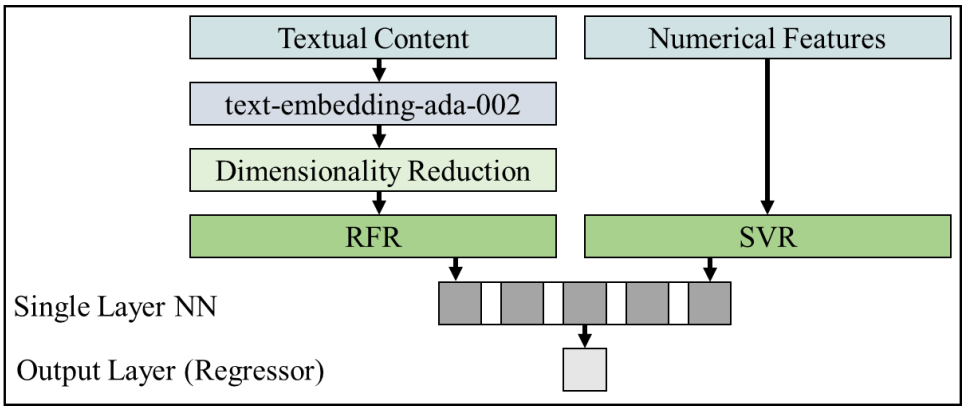

Figure 1. Models Architecture comprising the individual and meta regressors.

4 Preliminary Results

For assessing the meta regressors' performance we are adhering to common literature used metrics for reporting performance in regression tasks (Botchkarev, 2019): Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Squared Error (MSE), R-squared (R^2), and Mean absolute percentage error (MAPE). In order to evaluate the MAE, we specify the range of the log-transformed dependent variable, which is scaled between 0 and 13.72.

MAE	0.879
RMSE	1.154
MSE	1.332
R^2	0.579
MAPE	0.224

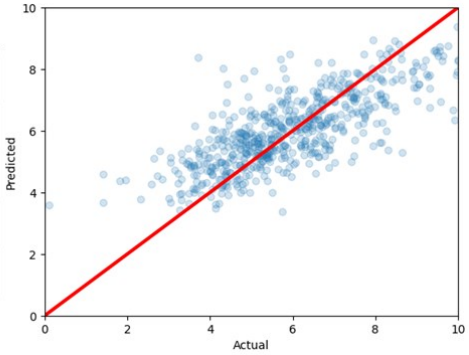

Figure 2. Relevant evaluation metrics and actual vs. predicted plot.

To have a base performance we trained an RFR solely on the numerical features resulting in an MAE of 1.210, an RMSE of 1.536, an MSE of 2.358, an R^2 of 0.270, and a MAPE of 0.225. Which is already close to the results achieved by the fined tuned RFR. In ensemble with the textual embedding features, the meta learning approach outperformed the base RFR with at least equal performance overall by far. The MAE for our meta-model was reported to be 0.879. Furthermore, the RMSE, a metric that gives higher weight to larger errors, was calculated to be 1.154. This suggests that while there are still some prediction errors, they are relatively infrequent. The MSE, which squares the errors before averaging, was 1.332. This indicates our regressor can avoid larger outliers successfully. The R^2 value, which measures the proportion of variance in the dependent variable was 0.579. This states that approximately

57.9% of the variance in the number of views is explained by our model, a substantial amount considering the complexity and variability inherent in online content engagement. The model's performance is summarized in Figure 2, where the distribution of errors and the fit of the model are illustrated. The actual versus predicted plot depicts a slight bias where the model tends to overpredict weaker presentations and underpredict stronger ones. This pattern indicates potential for future improvements in our model to further tackle data imbalance and features. Data imbalance has already been addressed to some extent with the log transformation and the exclusion criteria.

5 Discussion and Implications

The results from our meta-model, particularly the MAPE of 0.224 and R-squared of 0.579, demonstrate its effectiveness in predicting the number of views for online presentations. This performance significantly surpasses that of an RFR baseline. Our approach is, to the best of our knowledge, the first to incorporate an approach without domain-specific features into the variables. This approach adds yet to existing literature (Luzardo et al., 2014; Kim et al., 2017; Yi et al., 2022) by offering greater flexibility since it eliminates the need to reconstruct or reweight variables when adapting the model to a different platform. The meta-model streamlines evaluation processes, reducing manual effort while integrating into content management platforms for real-time insights and post-upload analysis. It offers two key contributions: automated presentation evaluation and accelerated evaluations leading to cost savings.

Our study, while providing valuable insights, is not without its limitations. Firstly, the performance of our model in predicting the number of views for presentations is contingent upon the dataset used. The model's performance is currently tailored to the inclusion criteria for data entries, which might limit its generalizability.

5.1 Next steps and experimental validation

To complete this research, the next steps are twofold. First, the diversity of our dataset across various domains illustrates the model's generalizability. However, domain-specific applications, particularly in research talks, promise further dual advantages: enhancing consistent feature availability, notably through the utilization of scholars' h-index, and introducing coverage in the reputational category. Additionally, we aim to explore DL models with multimodal capabilities for image analysis and talk transcription. These features could significantly further improve model performance.

Lastly, to validate the model's effectiveness, an experimental setting similar to the platform's slide deck evaluation process should be conducted. It should be assessed whether the model can aid and reduce the cognitive load in slide deck evaluation or if it introduces new complexities. At least 100 participants, content creators, should be equally split into two groups: one using the model as a DSS and a control group without it. Participants assess each presentation in all IQ categories and then categorize it as high, low, or moderate in potential viewer audience reach.

Acknowledgment

This work was supported by Innosuisse and is part of the flagship project SCESC, PFFS-21-29.

References

Awad, M. and Khanna, R. (2015). "Support Vector Regression." in: M. Awad & R. Khanna (Eds.), *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers* (pp. 67–80). Berkeley, CA: Apress.

Balikas, G. (2023). "Comparative Analysis of Open Source and Commercial Embedding Models for Question Answering." in: *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management* (pp. 5232–5233). New York, USA: ACM.

Botchkarev, A. (2019). "Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology." *Interdisciplinary Journal of Information, Knowledge, and Management*, *14*, 045–076.

Cagliero, L. and Quatra, M. L. (2021). "Automatic slides generation in the absence of training data." in: *2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 103–108). Madrid, Spain: IEEE.

Deng, L. and Liu, Y. (Eds.). (2018). *Deep Learning in Natural Language Processing*. Springer.

Drapkina, O. M., Volkova, I. Yu., Shepel, R. N., Zhamalov, L. M., Astanina, S. Yu. and Vakhovskaya, T. V. (2022). "Analysis of the quality of education activities conducted using telemedicine technologies." *Cardiovascular Therapy and Prevention*, *21*(3S), 3324.

Fenniak, M., Stamy, M., pubpub-zz, Thoma, M., Peveler, M., exiledkingcc, and pypdf Contributors. (2022). "The pypdf library."

Fu, T.-J., Wang, W. Y., McDuff, D. and Song, Y. (2022). "DOC2PPT: automatic presentation slides generation from scientific documents." in: *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, pp. 634–642).

Guo, Y., Zhang, Z., Liang, Y., Zhao, D. and Duan, N. (2023, November 7). "PPTC Benchmark: Evaluating Large Language Models for PowerPoint Task Completion." arXiv.

Ha, A. (2020, August 11). "Scribd acquires SlideShare from LinkedIn."

Hosseini, Z., Hytönen, K. and Kinnunen, J. (2022). "Improving Online Content Quality Through Technological Pedagogical Content Design (TPCD)," 284–296. Presented at the The Fourth Annual International Symposium "Education and City: Quality Education for Modern Cities."

Kim, S., Lee, J.-G. and Yi, M. Y. (2017). "Developing information quality assessment framework of presentation slides." *Journal of Information Science*, *43*(6), 742–768.

Kotsiantis, S. B., Zaharakis, I. D. and Pintelas, P. E. (2006). "Machine learning: a review of classification and combining techniques." *Artificial Intelligence Review*, *26*(3), 159–190.

Kühl, N., Hirt, R., Baier, L., Schmitz, B. and Satzger, G. (2021). "How to Conduct Rigorous Supervised Machine Learning in Information Systems Research: The Supervised Machine Learning Report Card." *Communications of the Association for Information Systems*, *48*(1).

Liu, Q., Kusner, M. J. and Blunsom, P. (2020, April 13). "A Survey on Contextual Embeddings." arXiv.

Luzardo, G., Guamán, B., Chiluiza, K., Castells, J. and Ochoa, X. (2014). "Estimation of Presentations Skills Based on Slides and Audio Features." in: *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*. Istanbul Turkey: ACM.

McInnes, L., Healy, J. and Melville, J. (2020, September 17). "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." arXiv.

Nocedal, J. and Wright, S. J. (Eds.). (2006). "Quasi-Newton Methods." in: *Numerical Optimization* (pp. 135–163). New York, NY: Springer.

polleverywhere. (2016). "10 little-known facts about PowerPoint." URL: https://blog.polleverywhere.com/powerpoint-infographic (visited on November 16, 2023)

Rhyn, M. and Blohm, I. (2017). "A Machine Learning Approach for Classifying Textual Data in Crowdsourcing."

Sefid, A., Wu, J., Mitra, P. and Giles, C. L. (2019). "Automatic Slide Generation for Scientific Papers."

Siddiqi, M. A. and Pak, W. (2021). "An Agile Approach to Identify Single and Hybrid Normalization for Enhancing Machine Learning-Based Network Intrusion Detection." *IEEE Access*, *9*, 13749.

Tekin, E., You, Q., Conathan, D. M., Fung, G. M. and Kneubuehl, T. S. (2022). "Harvest – a System for Creating Structured Rate Filing Data from Filing PDFs." in: *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 36, pp. 12414–12422).

Thelwall, M. and Kousha, K. (2016). "SlideShare Presentations, Citations, Users and Trends: A Professional Site with Academic and Educational Uses." *Journal of the Association for Information Science and Technology*, *68*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Polosukhin, I. (2023, August 1). "Attention Is All You Need." arXiv.

Yi, S., Matsugami, J. and Yamasaki, T. (2022). "Assessment system of presentation slide design using visual and structural features." *IEICE TRANSACTIONS on Information and Systems*, *105*(3), 5.

Zhang, C., Bengio, S., Hardt, M., Recht, B. and Vinyals, O. (2017, February 26). "Understanding deep learning requires rethinking generalization." arXiv.