A Comparative Analysis of MOOC Platforms Using Educational Data Mining Techniques

K. S. Savita[1], Pradeep Isawasan[2], Muhammad Akmal Hakim Ahmad Asmawi[2,*], Muhammad Shaheen[3], and Rabiya Ghafoor[4]

[1]Department of Computing, Positive Computing Center, Universiti Teknologi PETRONAS, Malaysia
[2]Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Perak Branch, Malaysia
[3]Faculty of Engineering and IT, Foundation University Islamabad, Pakistan
[4]Department of Software Engineering, Foundation University Islamabad, Pakistan
Email: savitasugathan@utp.edu.my (K.S.S.); pradeep@uitm.edu.my (P.I.); 2024655976@student.uitm.edu.my (M.A.H.A.A.); dr.shaheen@fui.edu.pk (M.S.); rabiya.ghafoor@fui.edu.pk (R.G.)
*Corresponding author
Manuscript received April 10, 2025; revised April 27, 2025; accepted May 28, 2025; published November 19, 2025

***Abstract*—Massive Open Online Courses (MOOCs) and microcredential have transformed education by providing flexible and accessible learning opportunities. This study conducts a comparative analysis of MOOC platforms, focusing on Coursera and Udemy, using Educational Data Mining (EDM) techniques. The research examines differences in course volume, subject trends, learner engagement, and course structure across platforms. The findings reveal that Coursera offers fewer but more structured and academically rigorous courses, often linked to university and corporate partnerships. Udemy, in contrast, provides a vast number of shorter, skill-focused courses with greater variability in quality due to its open-marketplace model. Coursera's courses have higher ratings and more consistent engagement, while Udemy follows a winner-takes-most pattern, where a few popular courses dominate learner interest. Text analysis of course titles highlight distinct subject preferences. Both platforms prioritize technology and business topics, but Coursera's content aligns with long-term career pathways, whereas Udemy quickly adapts to emerging industry trends, such as blockchain and artificial intelligence. The study also identifies differences in learning pathways, with Coursera structuring its content into beginner, intermediate, and advanced levels, while Udemy heavily relies on "all-level" courses to attract a broad audience. These insights provide valuable implications for educators, policymakers, and online learning providers. Structured platforms like Coursera can strengthen their credibility through academic rigor and industry partnerships, while market-driven platforms like Udemy can enhance content discovery and quality control. Future research should explore the role of course pricing, instructor expertise, and microcredential effectiveness in learner success.**

***Keywords*—Massive Open Online Courses (MOOCs), microcredential, educational data mining, Coursera, Udemy, text analysis, online education**

I. Introduction

Massive Open Online Courses (MOOCs) and microcredential have significantly transformed the educational landscape by offering accessible, flexible, and cost-effective learning opportunities to a global audience [1, 2]. As adoption increases among higher education institutions, corporate training providers, and independent instructors, these digital learning models have become central to skill development and professional certification [3]. Platforms such as Coursera, EdX, Udemy, and Alison offer a wide range of courses spanning technical, business, and academic subjects [4]. These platforms differ substantially in their operational models, ranging from institution-led formats (e.g., Coursera, EdX) to open marketplaces (e.g., Udemy, Alison) which in turn influence course structure, pricing, learner engagement, and accessibility [5].

To analyze the vast and diverse data generated by these platforms, Educational Data Mining (EDM) has emerged as a powerful approach [6]. Using techniques such as clustering, statistical modeling, and text analysis, EDM studies explore learner behaviors, course completion trends, and the effectiveness of microcredential pathways [7]. Common EDM themes include predicting student performance, identifying at-risk learners, and evaluating instructional designs which are the crucial factors to the future of digital education [8]. However, many existing studies focus on single platforms or limited metrics, often neglecting how structural differences across platforms shape user behavior and learning outcomes [9, 10].

While MOOCs have been widely studied, cross-platform comparisons that combine structural and behavioral insights remain limited [1, 11]. In particular, differences in course creation processes, subject emphasis, engagement patterns, and integration of microcredential are underexplored. Moreover, the lack of a standardized framework for comparing platforms has given variations in course length, pricing, and certification structures which complicates efforts to assess platform effectiveness in different educational contexts [1]. To address these gaps, this study conducts a comparative analysis of several MOOC platforms, with a focused case study on Coursera and Udemy. The objectives are as follows:

- To analyze course volume and identify popular subjects using text analysis across different MOOC platforms
- To examine ratings, reviews, course durations, and levels to uncover patterns in learner engagement.
- To compare Coursera and Udemy in depth, focusing on differences in content structure, platform strategy, and microcredential relevance.

This comparative approach deepens our understanding of how platform design and data-driven insights influence learning opportunities and outcomes in the MOOC ecosystem. By evaluating both institution-led and marketplace-driven models, the study highlights how platforms cater to diverse learner needs and career goals. These findings offer practical insights for educators, institutions, and policymakers seeking to optimize online education strategies. Additionally, the

exploration of subject trends and engagement metrics provides a foundation for understanding content effectiveness, learner participation, and the evolving role of microcredential in digital education.

II. Literature Review

Massive Open Online Courses (MOOCs) have transformed online education by offering open enrolment and flexible learning at scale [12]. Initially popularized by platforms like Coursera and edX, MOOCs now serve diverse learners, from college students to professionals seeking upskilling opportunities, supporting global trends of lifelong learning [13]. While early interest was driven by the promise of democratizing education, MOOC providers have since adopted varying models to balance financial sustainability with accessibility. Institution-led platforms (e.g., Coursera, edX) partner with universities to deliver accredited microcredential and certificates for a fee [11, 14], whereas marketplace platforms (e.g., Udemy, Alison) allow individual instructors to monetize content across academic and non-academic domains [15, 16]. These models differ in production quality, instructor autonomy, and pricing strategies, shaping learner engagement metrics like enrolment, ratings, and completion. However, high enrolment does not guarantee meaningful outcomes, as many users register out of curiosity, leading to high attrition [9, 17].

Educational Data Mining (EDM) plays a critical role in understanding learner behavior in such digital settings. EDM employs computational and statistical methods to analyze large-scale learning data, making it essential in MOOCs for identifying patterns, predicting outcomes, and improving design [18]. Key EDM applications include predictive modeling (to identify at-risk learners), clustering and segmentation (to group learners by behavior or demographics), and text mining (to analyze feedback or sentiment in discussion forums, reviews, and social media) [19]. These insights support adaptive learning recommendations and inform course improvements, leading to higher satisfaction and outcomes.

MOOC subject offerings align with evolving workforce needs, with popular areas including programming, data science, AI, and business management [20]. Emerging topics like blockchain, cybersecurity, and large language models (e.g., ChatGPT) have prompted more specialized curricula. Text mining techniques are increasingly used to track these trends through course titles and reviews. Marketplace platforms often respond more rapidly to such shifts due to flexible course-creation models, while institution-led platforms require more time for content development. Comparative studies highlight platform-specific strengths, Coursera's structured academic collaborations vs. Udemy's breadth and instructor freedom and show how engagement metrics vary by course type and platform features [1].

Despite growing interest, current research is limited by narrow scopes and methodologies, often focusing on single platforms or small datasets. Few studies apply robust EDM techniques across multiple platforms, limiting generalizability and missing broader behavioral patterns [1]. Additionally, the influence of MOOC business models on learner motivation and outcomes remains underexplored [21]. Addressing these gaps through structured, data-driven cross-platform analysis can provide deeper insights into what makes MOOCs effective. Such research informs best practices for educators, platforms, and policymakers seeking to optimize online education.

III. Materials and Methods

This study adopts a methodology adapted from the Data Science Trajectories (DST) model [22] which provides a structured framework for conducting data-driven research. The model has been customized to align with the study's objectives by incorporating five key phases as illustrated in Fig. 1. This structured approach ensures a systematic analysis of MOOC platforms, course engagement trends, and subject distributions. The Business Understanding phase, informed by the Related Work, involves defining the research objectives, establishing the scope of analysis, and identifying relevant MOOC platforms for comparison. In the Data Acquisition phase, course datasets are collected from an open source platform, Kaggle containing multiple MOOC providers, including Udemy, Coursera, EdX, and others, ensuring a diverse dataset for comparative analysis.

The Data Preparation phase includes cleaning, preprocessing, and standardizing course attributes, handling missing values, and preparing textual data for subject trend analysis. In Modeling, analytical techniques such as text analysis for keyword extraction, engagement metrics evaluation, and comparative studies of platform structures are applied to derive meaningful insights. Finally, Result Exploration involves interpreting findings, identifying key trends in MOOC platform strategies, subject preferences, and learner engagement, and presenting visualizations that highlight these insights. The Modeling and Result Exploration phases will be further detailed in the Results and Discussion section, where the study's findings will be examined in depth. This structured methodology ensures a comprehensive, data-driven approach to understanding MOOC platform effectiveness, subject demand, and engagement patterns, providing valuable insights for educators, policymakers, and industry stakeholders.

Fig. 1. Research methodology.

A. Data Acquisition

The Data Acquisition phase involves collecting MOOC course datasets from Kaggle, covering 10 MOOC providers with varying levels of data availability. The platforms included in this study are Alison, Coursera, London School of Economics (LSE), Berkeley, EdX, MIT OpenCourseWare (MIT OCW), Harvard University, Oxford University, Stanford University, and Udemy. These datasets provide essential attributes such as course titles, categories, ratings, reviews, course type, and pricing models, although not all platforms contain the same set of variables. The datasets were collected in CSV format from Kaggle between January and February 2025. Each file was manually reviewed to ensure key variables such as course title, ratings, and reviews were available. Datasets with missing schema or limited entries were excluded. The metadata (last update date, source verification, and number of records) was also recorded for

reproducibility. Due to the differences in available variables across datasets, some analyses are restricted to only platforms with common attributes. For example, engagement metrics such as review counts and enrolment numbers are not consistently available, limiting comparative engagement analysis. Similarly, pricing data is missing from certain platforms, affecting the ability to analyze cost-based trends. To mitigate these inconsistencies, the study focuses on universally available variables, ensuring a fair and transparent comparison across MOOC providers. The datasets were selected based on their coverage of major MOOC platforms, completeness of key attributes, and relevance to current online learning trends. After collection, the data was structured and standardized, ensuring that course categories, engagement metrics, and platform-specific attributes were aligned for further analysis. The datasets represent course information up to late 2024. Although this provides a near-current snapshot of MOOC platforms, the rapidly evolving nature of online education suggests that future work should include dynamic or API-based data to capture real-time platform updates. An overview of the dataset characteristics is shown in Table 1, highlighting the differences in available variables across platforms.

Table 1. The availability of variables within each datasets

Platform	No. of courses	Categories Availability	Ratings and No. of Reviews	Course Type	Pricing Info	Other Notable Features
Alison	4,940	Yes	No	Yes	No	Duration of courses, No. of learners, and Skills
Coursera	986	No	Yes	Yes	No	Level
London School of Economics	41	No	No	Yes	No	Mode, Duration and Category
Berkeley	45	No	No	Yes	No	Institute and Mode (Online/Face-to-Face/Hybrid)
EdX	1,278	No	No	Yes	No	Institute
MIT OCW	2,053	Yes	No	Yes	No	Subject, Resource type and Category
Harvard University	418	Yes	No	No	Yes	Duration, Mode and Subject
Oxford	834	No	No	No	No	Mode and Duration
Stanford	560	No	No	No	No	Institute and Mode
Udemy	25,443	No	Yes	No	No	Duration, and Level

B. Data Preparation

To ensure consistency and usability across the collected MOOC datasets, several data preprocessing steps were performed, addressing format inconsistencies, missing values, and duplicate records. Given that the datasets originated from different platforms with varying structures, a standardized approach was necessary before conducting further analysis. The first step involved formatting and structuring the data to create a uniform dataset suitable for comparison. All CSV files were loaded using Pandas in Python. Schema mapping was done to align inconsistent column names (e.g., "course_title", "Title"). Text fields were normalized by lowercasing and stripping special characters. Numerical values such as ratings and reviews were converted to float and cleaned of null or invalid entries. No imputation was applied to missing values to avoid introducing bias. Since column names and variable formats differed across sources, all attributes were renamed to follow a consistent naming convention. For example, fields such as "*Course_Title*" and "*Title*" were standardized under a single attribute name to maintain uniformity. Additionally, categorical variables such as course types were encoded into numerical formats where necessary, while numerical fields such as ratings and review counts were standardized to a uniform scale. This process ensured that data from different platforms could be merged seamlessly, maintaining alignment across attributes for effective comparative analysis.

Following the structuring process, missing data handling was carried out to address gaps in the datasets. Since certain platforms did not provide complete records for attributes like ratings, enrolments, or pricing, a selective approach was applied. Records missing essential fields, such as course titles or categories, were removed to maintain data reliability. However, other missing values were left as they were, ensuring that no artificial bias was introduced through imputation. Instead of estimating missing numerical values, the study worked with the available information, acknowledging these limitations in the analysis. The final step in the data preparation phase involved removing duplicate records to prevent redundancy. Since some courses appeared multiple times across different datasets, basic Python operations were used to detect and eliminate duplicates. Exact matching on course titles was performed using the "*drop_duplicates()*" function in pandas, ensuring that identical course records were removed while retaining only the first occurrence. This method effectively eliminated redundant listings, preventing duplicate courses from skewing engagement analysis or subject trend evaluations.

C. Modelling

The Modeling phase applied analytical techniques to compare MOOC platform structures, course distributions, and engagement patterns, providing insights into course availability, learner preferences, and subject trends. The analysis was conducted in two stages: a general comparative analysis across all platforms, followed by a more detailed investigation of Coursera and Udemy. The first stage focused on a comparative analysis of the number of courses offered by each MOOC platform. This provided an overview of platform differences in course volume, distinguishing between institution-led platforms and marketplace-driven platforms. Additionally, a word cloud analysis of course titles across all platforms was performed to identify popular course topics, highlighting common themes in online education.

The second stage involved a deeper analysis of Coursera and Udemy, two of the most prominent MOOC providers with differing business models. The analysis began with

summary statistics for both datasets, examining the distribution of key course attributes, including ratings, reviews, duration, and course levels. Data visualization and statistical analysis were performed using Python, specifically with the Pandas, Matplotlib, and WordCloud libraries. Summary statistics were calculated using built-in functions (mean, median, minimum, and maximum), and visual distributions were plotted as histograms and bar charts. Word cloud visualizations were generated by tokenizing course titles, removing stopwords, and rendering the most frequent keywords using the WordCloud library with frequency thresholds to highlight dominant themes. This allowed for a comparison of how courses are structured and received by learners on each platform. Further, distribution analyses were conducted for Coursera's ratings, reviews, course duration, and course levels, providing insights into how learners engage with different course formats. The study then used word cloud analysis on Coursera and Udemy course titles separately to identify which course topics are most popular on each platform. This helped reveal differences in content trends, where Coursera's offerings focus more on structured academic and professional development courses, while Udemy's content is more skill-based and influenced by emerging industry trends.

This modeling approach provides a comprehensive evaluation of MOOC platform structures, course engagement patterns, and subject demand, offering insights into how different learning platforms cater to various audiences and industry needs. The next section presents the findings and interpretations derived from these analyses.

IV. Result and Discussion

A. Overview of the Datasets

The data provides insights into the relative volume of course offerings across various prominent online learning platforms (see Fig. 2). The numbers associated with each platform denote the variety and possibly the depth of educational content they provide, which directly correlates to their market reach and influence within the rapidly expanding field of Massive Open Online Courses (MOOCs) and microcredential.

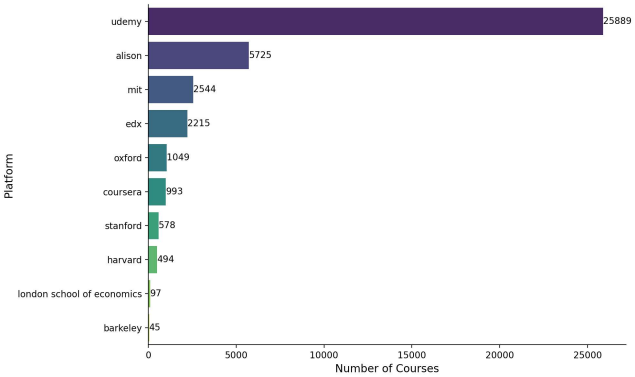

Fig. 2. Number of courses in each platform.

Udemy leads overwhelmingly, offering 25,443 courses which is more than all other platforms combined, it likely driven by its open marketplace model where anyone can create courses. While this vast selection provides accessibility to a wide range of niche topics, course quality can vary considerably, prompting learners to depend heavily on reviews and ratings. Alison and MIT also offer substantial course volumes, with 4,940 and 2,053 courses, respectively. Alison's focus on free, foundational, and skill-based courses appeals broadly to global learners seeking basic proficiency. MIT's extensive offerings through MIT OpenCourseWare underscore its institutional commitment to open education and knowledge sharing. In contrast, curated platforms like Coursera (986 courses) and edX (1,278 courses) prioritize partnerships with prestigious universities and industry leaders, offering fewer courses of typically higher quality. Coursera further distinguishes itself by providing structured pathways like microcredential and accredited online degrees, aligning closely with traditional higher education standards and enabling credit transfers to universities. Prestigious institutions such as Harvard, Oxford, and Stanford provide fewer but strategically selective offerings, leveraging their strong reputations to attract learners who value institutional prestige as much as course content. Platforms like Berkeley and the London School of Economics present minimal course counts (45 and 41, respectively), suggesting either recent entry into the digital education space or a targeted approach emphasizing highly specialized content aligning with their institutional strengths. This diversity illustrates the evolving MOOC landscape, shaped by different platform strategies, from open markets to carefully curated selections. As digital badges and microcredential increasingly gain industry acceptance, these platforms play an essential role in lifelong learning and professional development, reshaping traditional perceptions of educational qualifications.

B. Word Cloud of Course Titles

The word cloud in Fig. 3 contains courses from 10 different learning platforms provides insights into the current trends and demands within the realms of online education, MOOCs, and microcredential. "Python" emerges as the most dominant keyword, underscoring the huge popularity and necessity for skills in this programming language across various fields such as data science, machine learning, web development, and automation. This reflects the growing emphasis placed on coding and software development skills in the job market, where employers are increasingly seeking candidates who can navigate and leverage the power of data and software.

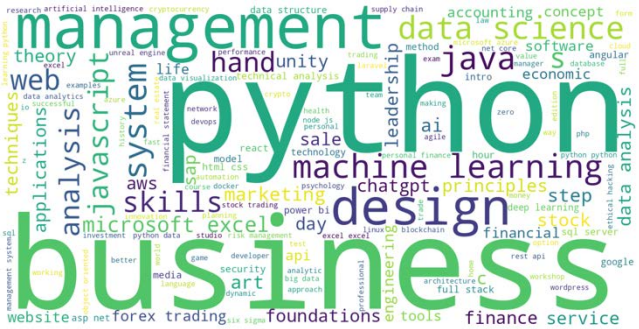

Fig. 3. Word cloud of course titles.

"Business" and "management" remain popular, signalling continued interest in leadership, strategy, and operational skills essential for organizational competitiveness. Related terms like "leadership," "team", and "successful" underline the rising importance of soft skills alongside technical competencies. High prominence of "design", "machine

learning", and "data science" illustrates an increasing fusion of creativity with technological advancement, particularly as industries adopt automation and AI-driven solutions. Niche keywords such as "Microsoft Excel", "AWS", "SAP", and "accounting" suggest learners seek targeted expertise in specific tools and systems, enhancing practical and functional skills that directly boost employability. Growing mentions of "marketing", "financial", "stock trading", and "forex trading" reflect heightened interest in financial literacy, investing, and wealth management. This trend is influenced by digital finance innovations like blockchain and cryptocurrency. Keywords such as "cloud", "Azure", and "DevOps" highlight the critical role of cloud computing and IT infrastructure in digital transformation initiatives. Additionally, the presence of "health", "healthcare", and "safety" points to increased awareness and educational interest influenced by recent global events, including the COVID-19 pandemic. Overall, these trends reveal clear connections between current market demands, technological progress, and learners' career aspirations, offering valuable guidance to both educators and students in shaping future skill development.

C. Summary of Statistics of Coursera Vs Udemy

Coursera and Udemy are two well-known online learning platforms that serve different types of learners and educational needs. A comparative analysis of their summary statistics in Table 2 reveals key differences in course ratings, user engagement, and course durations, showing their different structures and target audiences.

Table 2. The summary statistics for Coursera and Udemy

Metric	Coursera	Udemy
Average Rating	4.68	4.32
Median Rating	4.70	4.40
Minimum Rating	2.80	1.00
Maximum Rating	5.00	5.00
Average Review Count	7,827.92	1,555.68
Median Review Count	1,900.00	288.00
Minimum Review Count	6.00	1.00
Maximum Review Count	268,600.00	486,391.00
Average Duration (hours)	452.57	6.26
Minimum Duration (hours)	2.00	0.05
Maximum Duration (hours)	960.00	189.00

Coursera's courses consistently receive higher ratings, with an average of 4.68 versus Udemy's 4.32. Coursera's minimum rating (2.80) is also notably higher than Udemy's (1.00), highlighting Coursera's structured, quality-driven approach stemming from university and institutional partnerships. In contrast, Udemy's open marketplace model results in wider quality variation, though both platforms have highly-rated courses reaching a maximum rating of 5.00. Engagement metrics further differentiate the platforms. Coursera courses average significantly more reviews (7,828) compared to Udemy (1,556). The median reviews per course reinforce this trend (1,900 for Coursera versus 288 for Udemy), indicating higher overall learner engagement per course on Coursera. Although Udemy hosts some exceptionally popular courses with review counts surpassing Coursera's top performers, most Udemy courses attract considerably less attention. This uneven engagement highlights Udemy's marketplace dynamics, where a few top courses dominate learner interest, unlike Coursera's more evenly distributed popularity. Course durations vary dramatically between the platforms. Coursera offers lengthy, in-depth courses averaging 452.57 h, reflecting structured university-like curricula or comprehensive certification programs. Conversely, Udemy's courses average just 6.26 h, with some as short as three minutes. This short format suits learners seeking rapid, skill-specific training rather than prolonged study. Overall, Coursera appeals primarily to learners pursuing structured, academically rigorous, career-oriented education with consistent quality, whereas Udemy attracts learners who value affordability, flexibility, and quick skill-building opportunities. These findings underscore the importance of aligning platform choice with specific educational objectives and learner preferences.

D. Distribution of Ratings, Reviews, and Duration for Coursera vs Udemy

A comparative analysis of course duration, ratings, and reviews for Coursera and Udemy in Figs. 4 and 5 reveals significant differences in course structures, user engagement patterns, and rating behaviors across both platforms. These distinctions highlight how different instructional models and learner expectations shape each platform's course offerings.

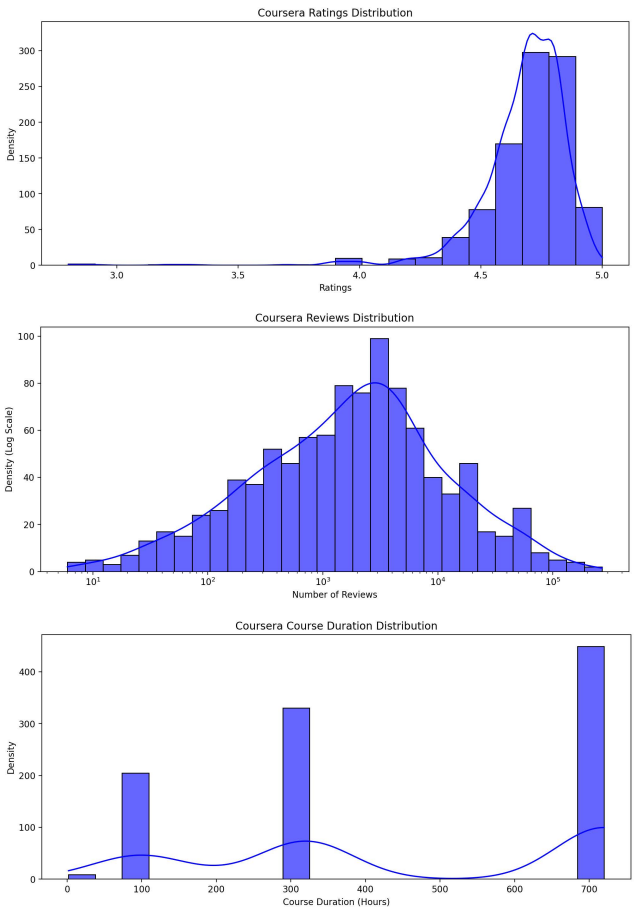

Fig. 4. Distributions of ratings, reviews and duration for Coursera.

Coursera exhibits a multi-peaked distribution in course duration, with clusters around 100, 300, and 700+ h. This pattern reflects its structured learning paths, ranging from short specializations to extensive professional certificates or university-level curricula. Shorter courses are often standalone modules, while longer durations correspond to comprehensive academic programs. In contrast, Udemy's

course duration is heavily skewed towards shorter courses, with the majority under 25 h. This aligns with its marketplace model, catering to learners seeking quick skill acquisition. Longer courses on Udemy are primarily found in technical and professional fields, where deeper study is required. Ratings also differ significantly between the platforms. Coursera maintains higher and more consistent ratings, with an average of 4.68 and a median of 4.70, suggesting a high-quality learning experience with strict academic oversight. Udemy, on the other hand, has a lower average rating of 4.32 with greater variability, ranging from as low as 1.00 to highly rated courses exceeding 4.5. This variability reflects Udemy's open-marketplace structure, where course quality can differ significantly based on instructor expertise and student expectations. Learners on Udemy rely more on reviews and social proof to assess course credibility, whereas Coursera ensures consistent standards through institutional partnerships.

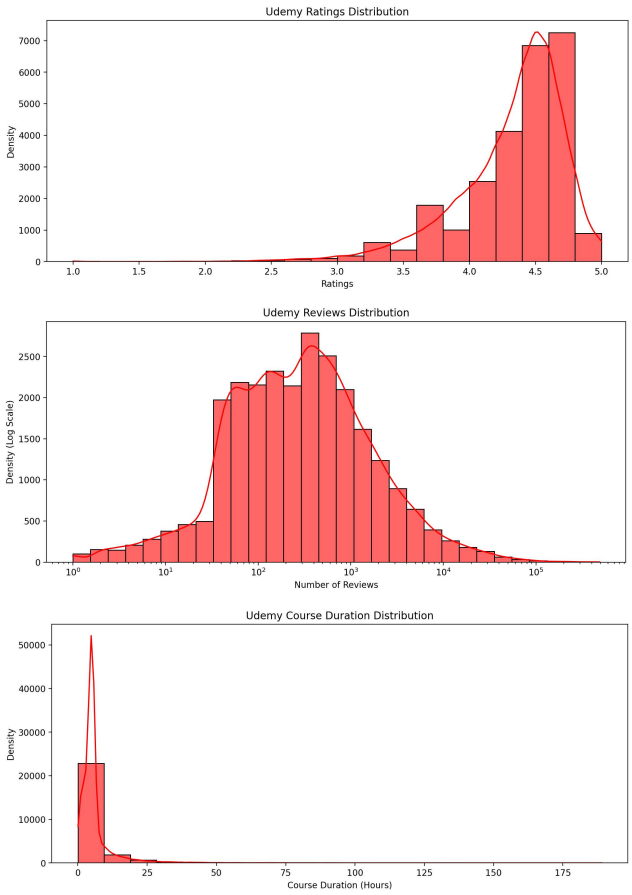

Fig. 5. Distributions of ratings, reviews and duration for Udemy.

Review distribution patterns further emphasize engagement differences. Coursera follows a long-tail effect, where most courses attract moderate engagement, with a few courses exceeding 100,000 reviews. This distribution suggests steady engagement across a variety of courses, as learners commit to structured educational pathways. Udemy, in contrast, follows a winner-takes-most pattern, where a small number of top courses garner massive engagement, with some courses amassing nearly 500,000 reviews. This disparity reflects Udemy's competitive marketplace, where top-rated courses dominate through platform promotion and learner-driven popularity, while many other courses receive significantly fewer reviews. Overall, these findings reinforce the distinct positioning of both platforms. Coursera's structured, high-quality environment attracts learners seeking academic rigor, career-aligned content, and recognized certifications. Its courses are longer, consistently well-rated, and evenly distributed in terms of engagement. Udemy, by contrast, functions as a dynamic marketplace with shorter, skill-based courses, highly variable ratings, and a sharp engagement divide between popular and lesser-known courses. While Coursera offers a more predictable and uniform learning experience, Udemy thrives on flexibility, affordability, and market-driven course popularity.

E. Distribution of Course Level for Coursera Vs Udemy

The distribution of course levels on Coursera and Udemy in Figs. 6 and 7 highlights key differences in content structure and target audience. While both platforms aim to provide accessible learning opportunities, their approach to course categorization reflects differing priorities in content curation and learner engagement.

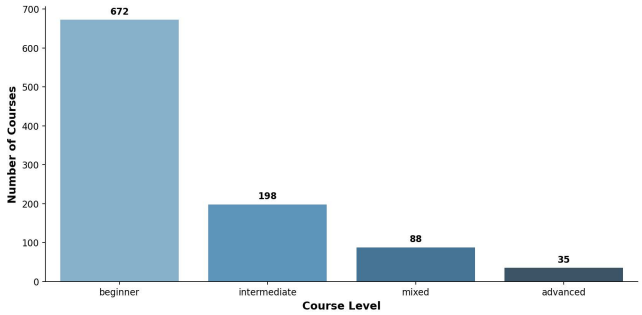

Fig. 6. Distributions of course level for Coursera.

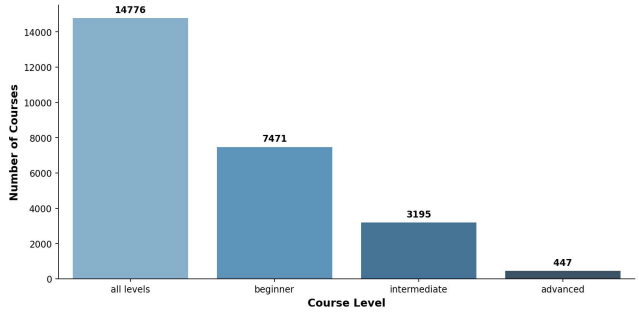

Fig. 7. Distributions of course level for Udemy.

Coursera primarily emphasizes beginner-level courses (715), catering to learners acquiring fundamental knowledge or exploring new fields. Its structured learning paths, such as specializations and professional certifications, guide users progressively, with 198 intermediate-level courses supporting advancement in areas such as data science, programming, and business. Fewer courses appear in advanced (25) and mixed "all levels" (68) categories, reinforcing Coursera's structured academic-driven progression from introductory to intermediate content. This distribution aligns with Coursera's partnerships with universities and industry leaders, ensuring learners follow a defined learning path toward expertise. In contrast, Udemy features a significantly higher number of courses labeled "all levels" (14,776), reflecting its open, flexible learning model designed for a wide range of learners. Although beginner courses (7,673) are abundant, Udemy also provides more intermediate (3,195) and advanced (447) courses than

Coursera, indicating a greater diversity of content and specialization opportunities. This distribution suggests that Udemy accommodates both learners seeking introductory knowledge and those looking for deeper, advanced skill development. Unlike Coursera, which guides learners through a structured pathway, Udemy's open-marketplace approach allows students to jump into courses at any level based on their personal preferences.

These differences in course level distribution illustrate how Coursera and Udemy serve distinct learner needs. Coursera excels in guiding learners through clear educational pathways, making it ideal for individuals looking for formalized learning structures and career certifications. Udemy, on the other hand, accommodates a broader audience by providing flexibility, making it a strong option for those seeking self-paced, skill-based learning at various proficiency levels. The dominance of beginner-level courses across both platforms suggests that online education remains largely focused on accessibility, helping learners enter new fields with ease. However, Udemy's larger selection of intermediate and advanced courses positions it as a more versatile platform for learners seeking specialized skills beyond foundational knowledge.

F. Word Cloud of Course Titles for Coursera vs Udemy

A comparative analysis of course titles from Coursera and Udemy in Figs. 8 and 9 provides valuable insights into their focus areas and target learners. The word clouds highlight distinct differences in topic emphasis, learning style, and industry alignment between the two platforms.

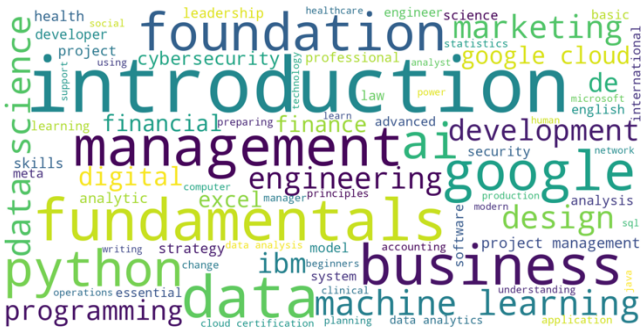

Fig. 8. Word cloud of course titles from Coursera.

Fig. 9. Word cloud of course titles from Udemy.

Coursera strongly emphasizes structured and foundational learning, as indicated by dominant keywords such as "introduction", "fundamentals", and "foundation". This suggests a deliberate focus on entry-level knowledge and progression from beginner to intermediate levels. Additionally, Coursera's frequent references to "data", "machine learning", and "management" reflect its career-oriented approach, particularly in technology and business fields. Industry-linked certifications are evident in Coursera's inclusion of terms such as "Google", "IBM", and "Microsoft", reinforcing its alignment with formal academic credentials and corporate partnerships. In contrast, Udemy's course titles indicate a practical, skills-based focus, with dominant keywords like "complete", "learn", "masterclass", and "bootcamp". This reflects Udemy's emphasis on comprehensive, hands-on learning experiences designed for quick skill acquisition. The platform covers a broad range of technical skills, including "Python", "Java", "SQL", "AWS", and "JavaScript", showcasing its strength in coding, cloud computing, and IT certifications. Additionally, Udemy is more responsive to emerging trends, as seen in frequent references to "ChatGPT", "blockchain", and "cryptocurrency", demonstrating its ability to quickly adapt course offerings to market demand.

Both platforms prominently feature technology-related keywords such as "Python", "machine learning", and "data science", underlining the growing demand for technical proficiency in the job market. However, Coursera's structured, academically aligned curriculum contrasts with Udemy's independent, skill-based approach. Business-related terms like "management", "marketing", and "finance" appear across both platforms, but Coursera integrates them into formal academic frameworks, while Udemy emphasizes practical business tools and certifications, such as "Excel" and "project management". These differences highlight Coursera's role as a structured, academically oriented platform that offers industry-backed certifications and clear learning pathways, making it ideal for learners seeking formalized career progression. Udemy, on the other hand, prioritizes flexibility, affordability, and rapid skill-building, making it a preferred choice for learners looking to quickly upskill in both established and emerging fields.

V. Actionable Insights

A. Course Volume, Subject Trends, and Market Adaptability

MOOC platforms adopt different strategies in content structuring. Structured platforms like Coursera should continue leveraging academic credibility and career-aligned pathways, ensuring their courses align with long-term professional growth. Marketplace-driven platforms like Udemy can maximize adaptability, focusing on rapidly evolving industry trends and skill-based learning. Given the dominance of technology-related subjects, MOOC providers should expand offerings in high-demand fields like AI, blockchain, and cloud computing, while maintaining a balance between foundational knowledge and specialized expertise.

B. Learner Engagement and Course Characteristics

Engagement patterns indicate that structured platforms benefit from consistent content quality, while open-market platforms need better content discovery mechanisms. Coursera's model of evenly distributed engagement ensures reliability, whereas Udemy's winner-takes-most effect creates visibility challenges for less popular courses. Improving content recommendation systems and instructor

quality control can enhance engagement for marketplace-based MOOCs, ensuring a broader distribution of learner interaction. Additionally, platforms should align course durations with learner expectations—structured, long-form learning for deep skill development and shorter, modular content for fast upskilling.

C. Platform Differences in Course Levels and Learning Pathways

Beyond platform-specific insights, this study has broader implications. For platform developers, understanding learner engagement patterns can inform content recommendation systems, adaptive learning features, and quality assurance mechanisms. Educators can better align course design with platform dynamics, ensuring relevance and learner retention. Policymakers and institutional stakeholders can use these findings to inform digital education frameworks, microcredential policies, and workforce development programs that depend on scalable, accessible learning models.

VI. Conclusion

This study revealed key differences in MOOC platform strategies and learner engagement patterns. Coursera leans toward fewer, academically structured courses with steady interaction and consistently high ratings, aligning with formal educational goals. In contrast, Udemy emphasizes flexible, shorter courses that respond quickly to market demand, though often with greater variability in quality and engagement. However, several limitations affect the generalizability of these findings. The data came from static Kaggle datasets (2022–2024) not officially released by the platforms, possibly missing newer courses or behavioral trends. Metadata gaps like missing pricing or engagement details, limited full comparison, and the absence of user-level data (e.g., course completion or dropout rates) constrained engagement analysis. Broader external factors such as economic conditions or platform algorithm changes were also outside the scope. Still, the role of microcredentials emerged as critical across both platforms, with increasing demand for credentials tied directly to employability. Future work should explore variables like instructor credentials, course pricing, and long-term career impact to deepen understanding of learner motivations and optimize platform strategies.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

K.S. Savita conceptualized the study and supervised the research. Pradeep Isawasan and Muhammad Akmal Hakim Ahmad Asmawi conducted the data analysis. Muhammad Shaheen contributed to the literature review. Rabbeya Ghafoor assisted in data preprocessing and visualization. All authors reviewed and approved the final manuscript.

Funding

We would like to acknowledge the grant funder, ICRF UTP-FUI [015ME0-384] for funding this research and article publications, as well to reviewers for the constructive feedback.

References

[1] J. A. Ruipérez-Valiente *et al.*, "Large scale analytics of global and regional MOOC providers: Differences in learners' demographics, preferences, and perceptions," *Comput Educ.*, vol. 180, 104426, Apr. 2022. doi: 10.1016/j.compedu.2021.104426

[2] H. Wit and P. G. Altbach, "Internationalization in higher education: global trends and recommendations for its future," *Policy Reviews in Higher Education*, vol. 5, no. 1, pp. 28–46, Jan. 2021. doi: 10.1080/23322969.2020.1820898

[3] R. Narayanaswamy, C. S. Albers, T. L. Knotts, and N. D. Albers, "Sustaining and reinforcing the perceived value of higher education: E-Learning with micro-credentials," *Sustainability*, vol. 16, no. 20, 8860, Oct. 2024. doi: 10.3390/su16208860

[4] S. Varadarajan, J. H. L. Koh, and B. K. Daniel, "A systematic review of the opportunities and challenges of micro-credentials for multiple stakeholders: Learners, employers, higher education institutions and government," *International Journal of Educational Technology in Higher Education*, vol. 20, no. 1, p. 13, Feb. 2023. doi: 10.1186/s41239-023-00381-x

[5] I. Irwanto, D. Wahyudiati, A. D. Saputro, and I. R. Lukman, "Massive Open Online Courses (MOOCs) in higher education: A bibliometric analysis (2012–2022)," *International Journal of Information and Education Technology*, vol. 13, no. 2, pp. 223–231, 2023. doi: 10.18178/ijiet.2023.13.2.1799

[6] R. L. Moore and J. Shin, "Using learner engagement patterns for predictive modeling in a MicroMasters program," *Distance Education*, pp. 1–21, Jan. 2025. doi: 10.1080/01587919.2024.2441252

[7] A. Ani and E. T. Khor, "Development and evaluation of predictive models for predicting students performance in MOOCs," *Educ. Inf. Technol (Dordr)*, vol. 29, no. 11, pp. 13905–13928, Aug. 2024. doi: 10.1007/s10639-023-12398-w

[8] O. H. T. Lu, A. Y. Q. Huang, and S. J. H. Yang, "Impact of teachers' grading policy on the identification of at-risk students in learning analytics," *Comput Educ*, vol. 163, 104109, Apr. 2021. doi: 10.1016/j.compedu.2020.104109

[9] J. Goopio and C. Cheung, "The MOOC dropout phenomenon and retention strategies," *Journal of Teaching in Travel & Tourism*, vol. 21, no. 2, pp. 177–197, Apr. 2021. doi: 10.1080/15313220.2020.1809050

[10] A. N. Alshammari, "Do learning patterns differ by gender in MOOCs? Exploring gender-based differences through learning analytics," *International Journal of Information and Education Technology*, vol. 15, no. 3, pp. 617–628, 2025. doi: 10.18178/ijiet.2025.15.3.2270

[11] Y.-M. Bideau and T. Kearns, "A European approach to micro-credentials for lifelong learning and employability," *J. Eur. CME*, vol. 11, no. 1, Dec. 2022. doi: 10.1080/21614083.2022.2147288

[12] S. Papadakis, "MOOCs 2012–2022: An overview," *Advances in Mobile Learning Educational Research*, vol. 3, no. 1, pp. 682–693, Mar. 2023. doi: 10.25082/AMLER.2023.01.017

[13] J. Castaño-Muñoz and M. Rodrigues, "Open to MOOCs? Evidence of their impact on labour market outcomes," *Comput Educ*, vol. 173, 104289, Nov. 2021. doi: 10.1016/j.compedu.2021.104289

[14] G. Tamoliune, R. Greenspon, M. Teresevicíene, A. Volungeviciene, E. Trepule, and E. Dauksiene, "Exploring the potential of micro-credentials: A systematic literature review," *Front Educ (Lausanne)*, vol. 7, Jan. 2023. doi: 10.3389/feduc.2022.1006811

[15] M. T. Cisel and D. Pontalier, "Knowledge marketplaces: An analysis of the influence of business models on instructors' motivations and strategies," *The International Review of Research in Open and Distributed Learning*, vol. 22, no. 3, pp. 142–158, Apr. 2021. doi: 10.19173/irrodl.v22i3.5459

[16] D. De Notaris, S. Canazza, C. Mariconda, and C. Paulon, "How to play a MOOC: Practices and simulation," *Entertain Comput*, vol. 37, 100395, Mar. 2021. doi: 10.1016/j.entcom.2020.100395

[17] M. D. B. Castro and G. M. Tumibay, "A literature review: Efficacy of online learning courses for higher education institution using meta-analysis," *Educ Inf Technol (Dordr)*, vol. 26, no. 2, pp. 1367–1385, Mar. 2021. doi: 10.1007/s10639-019-10027-z

[18] H. A. Althibyani, "Predicting student success in MOOCs: A comprehensive analysis using machine learning models," *PeerJ Comput. Sci.*, vol. 10, e2221, Aug. 2024. doi: 10.7717/peerj-cs.2221

[19] M. Yee *et al.*, "AI-assisted analysis of content, structure, and sentiment in MOOC discussion forums," *Front Educ (Lausanne)*, vol. 8, Sep. 2023. doi: 10.3389/feduc.2023.1250846

[20] D. Shah. The second year of the MOOC: A review of MOOC stats and trends in 2020. [Online]. Available: https://www.classcentral.com/report/the-second-year-of-the-mooc/#:~:text=Now%20in%20its%20ninth%20year%2C,microcredentials%2C%20and%2019%20online%20degrees

[21] D. J. Camacho and J. M. Legare, "Pivoting to online learning—The future of learning and work," *The Journal of Competency-Based Education*, vol. 6, no. 1, Mar. 2021. doi: 10.1002/cbe2.1239

[22] F. Martinez-Plumed *et al.*, "CRISP-DM twenty years later: From data mining processes to data science trajectories," *IEEE Trans Knowl Data Eng*, vol. 33, no. 8, pp. 3048–3061, Aug. 2021. doi: 10.1109/TKDE.2019.2962680

Copyright © 2025 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited (CC BY 4.0).