MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 2282-2285

Data science competition platform for promotes solution sharing in learning sector

* Ghaliya Al-Farsi[1]
galfarsi@buc.edu.om
Al Buraimi University College

Abstract. The ongoing M4 competition exemplifies the significance of competition in the realm of forecasting. Scholars and industry experts have been drawn to the competition, sparking discussions regarding the data's adequacy for business forecasting. Nevertheless, numerous Kaggle competitions that present real-world forecasting issues have largely been overlooked by the academic sector. This paper outlines the findings from six Kaggle competitions, as it is believed that the forecasting community could significantly gain insights from these events. It is noted that the majority of Kaggle data sets exhibit greater variability and volatility compared to the M-competition. Furthermore, it is evident that gradient-boosted decision trees yield strong performance, demonstrating improvement in results.

Keywords— Platform Kaggle, Machine learning, artificial intelligence, technologies.

1 Introduction

Kaggle allows data scientists and machine learning practitioners to compete against each other on the platform, striving to develop superior models for specific challenges or analyzing particular datasets. The platform also provides a community space for users to share code and datasets, collaborate on projects, and learn from each other's contributions. In 2017, Google acquired Kaggle, and it is now integrated into Google Cloud services. Kaggle was founded in the year 2010. [1]. Organizations support a wide extend of competitions on Kaggle from picture classification to extortion location to therapeutic result expectation In expansion to seeing their execution on a open leaderboard and getting feedback from other competitors and the community members can yield their models [1] [2]. To help clients in learning and sharpening their machine learning and information science abilities Kaggle gives open information sets machine learning note pads and instructional exercises in expansion to competitions It is presently a wellliked stage for information researchers of all encounter levels to progress their information grow their portfolios and network with other experts within the field [3] [4][11][12][15].

Over time Kaggle has created relentlessly growing its offerings to incorporate a cloudbased environment of administrations that offer assistance with competitions It started by facilitating a parcel of datasets created into a open information stage andmost intriguinglystarted advertising its clients a webbased environment for information science that was driven by cuttingedge containerization innovation .

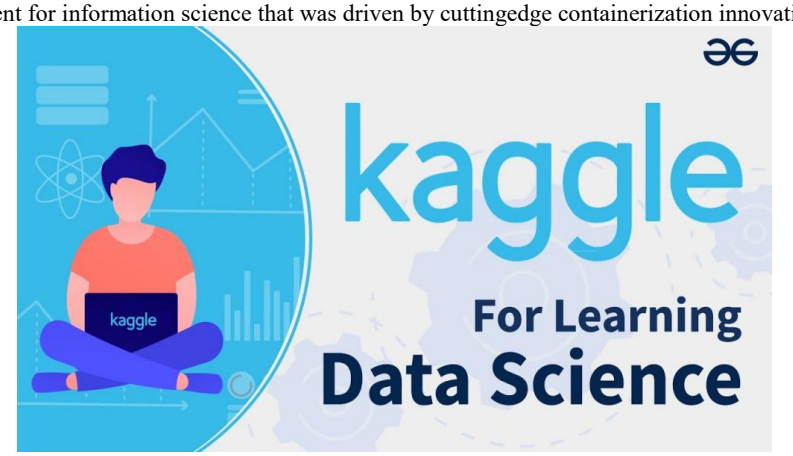

Fig. 1. Kaggle for learning Data scince

Not at all like the scholastic estimating competitions Kaggles accentuation on advertising arrangements for realworld determining assignments incorporates a downside in that it places confinements on the conclusions that scholastics can draw from the challenges It is outlandish to test for calculable varieties within the exhibitions of arrangements or to evaluate execution utilizing distinctive blunder measurements on the off chance that the test set is inaccessible after the competition has concluded [5] [6].Besides in order to superior appreciate the focal points and impediments of elective approaches it isn't practical to compare the execution of different arrangements on different subsets of the dataset Whereas Kaggle advances arrangement sharing members are not committed to form their code or arrangement freely accessible and the nonattendance of freely accessible arrangements has results.Anyone interested in data, Kaggle has become a vital resource. Whether you're a beginner or an expert in data science, follow these steps to use Kaggle efficiently for your research:

a) Create a Kaggle account: In order to begin, you must first create an account on Kaggle. Using your Google account or email, you can accomplish this.

b) Use Kaggle's interface: Get acquainted with the platform's intuitive design. Investigating the webpage, datasets, contests, and educational materials are all part of this.

c) Explore Kaggle datasets: Begin by browsing through Kaggle's extensive collection of datasets. Utilize the search features and filters to find datasets that align with your interests and projects.

d) Participate in Kaggle competitions: Choose a competition that matches your skill level.

Consider using Kaggle due to the following notable advantages:

1. All available datasets are provided transparently, and the challenges are clearly defined.
2. The tough nature of the public and private leaderboards makes it hard to deceive oneself with a poor testing setup.
3. Each competition is often accompanied by engaging discussions and sharing, which you can both contribute to and learn from.
4. You can curate a portfolio that showcases your skills by working with complex real-world datasets.
5. It operates as a pure meritocracy, where success is determined by talent and skill, rather than academic credentials or mathematical ability.

Is Engaging with Kaggle a Valuable Investment for Data Scientists?

Kaggle is a prominent platform that provides users with access to educational accelerators, allows exploration and publication of datasets, and facilitates participation in predictive modeling competitions. This vibrant ecosystem enables data scientists to develop remarkable machine learning models with the assistance of their peers.[20].The AI makes collaborative learning environments possible with Kaggle and the technology learning such as virtual learning assistants and intelligent tutoring systems can advise students, facilitate group discussions, and encourage teamwork. By encouraging active engagement, critical thinking, and problem-solving abilities, these AI-powered technologies provide a dynamic learning environment that simulates real-world situations [12] [13] [14].

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 2282-2285

2 Applications you used

Most data science such as python and Tensor Flow are just For anyone engaged in data, Kaggle has turned into an essential tool. Whether you are a novice or a seasoned data scientist, follow these steps to capitalize on Kaggle efficiently for your research purposes: a) Set up a Kaggle account: To start, you need to register for an account on Kaggle. This can be done using your Google account or your email address.are merely a handful of the widely recognized data science tools and packages that Kaggle presents. Engaging with Kaggle allows professionals to remain informed about the latest progressions in the field and gain hands-on experience with these technologies. A range of pre-trained models is accessible for various uses such as sentiment analysis, object detection, segmentation, and classification. If it piques your interest, Kaggle offers a space for researching datasets. [7] [8][9].

3 Literature review

It alludes to the method of consequently assessing and summarizing inquire about papers and entries relating to machine learning and information science within the setting of Kaggle an AI powered stage. This keeps experts up to date on the foremost later progressions and revelations in their industry More details are accessible on Kaggles stage on the off chance that youre curious about exploring [10] [13] [14].
AI powered writing surveys Januschowski Concurring to Januschowski et al (2022) [16] [17]. The qualification is ineffectual since it comes up short to offer an exact and unbiased classification of strategies For occurrence the best two approaches within the M4 competition combine factual procedures like exponential smoothing and other conventional time arrangement determining strategies with ML methods like neural systems and slope boosted choice trees GBDT. Januschowski et al 2020, give an interchange approach that includes utilizing a broader extend of classification measurements such as worldwide versus nearby models straight versus nonlinear information ambitious versus classic ambitious outfit versus lone prototypes with interpretable versus prescient in it. Kang Putting the found the cooperation datasets popular connection to the M3 and M4 competitions is point of the examination handle can gather that the datasets utilized in Kaggle challenges are characteristic of particular real world scenarios since they appear the genuine estimating issue of well-known organizations 2017 to speak to a single time arrangement in 2 ways interplanetary empowering the ponder of important period arrangement datasets and contextualize the Kaggle competition concerning the M4 and M3 datasets. It allows for discussions regarding the results of M4 and M3 competitions in contexts that resemble those of the mentioned Kaggle contests [16] [17]. Makridakis The M4 competition incorporated higher frequency data in the processing of weekly, daily, and hourly datasets to address prediction uncertainty, requesting prediction intervals. To alleviate concerns regarding the statistical relevance of the findings, the sample size was expanded to 100,000 time series, a focus on repeatability was prioritized, and various established methods were employed as benchmarks. Only continuous time series were allowed, most of which stemmed from the business sector. Furthermore, the time series were required not to be intermittent or to contain gaps (Spiliotis et al., 2020). However, the weekly time series only needed 80 observations (Makridakis et al., 2020b), with each frequency necessitating more than three complete seasonal cycles [18]. Darin and Stellwagen (2020) and Fry and Brundage (2019): Some professionals have raised concerns about the representativeness of the dataset used in the competition, questioning the relevance of the M4 competition results to business forecasting within Kaggle. According to Darin and Stellwagen (2020) and Fry and Brundage (2019) [19], numerous forecasting challenges that businesses encounter are present. Let's break down the various methods and procedures employed in Kaggle's initial data analysis.

data visualization: Creating graphs (such as scatter plots, histograms, etc.) to explore relationships and trends.
3- addressing missing data: Recognizing and managing absent values.
4- spotting outliers: Identifying extreme values.
5- crafting new features: Generating additional features from existing ones.
6- analyzing correlation: Understanding relationships across variables.
7- Initial Analysis of Application Dataset: At this phase, a specific dataset is evaluated initially.
While the particulars may vary, it generally includes:
1- Data Overview: Assessing the scope, structure, and essential statistics.
2- Data cleansing: Handling outliers, duplicates, and absent values.
3- visualization: Graphing data distributions to observe their characteristics.
Information Choice on Kaggle Kaggle gives a wide extend of determining datasets frequently sourced from real world applications industry accomplices or engineered datasets outlined to test machine learning models. The information choice handle regularly takes after these steps: Dataset Accessibility Kaggle competitions regularly highlight datasets given by organizations or scholarly analysts.
Preprocessing Needs Datasets experience cleaning lost esteem ascription and include designing some time recently investigation.
Time Series Characteristics Information incorporates timestamps regularity impacts and outside factors.
Assessment Criteria Common mistake measurements such as RMSE MAE and MASE are predefined to benchmark execution.

4 Result and Discussion

Comes about on Kaggle ordinarily relate to the results of machine learning errands or information science competitions These results are as often as possible shown as leaderboard rankings where members models are surveyed agreeing to foreordained criteria eg exactness F1 score or cruel supreme mistake.There are talks tending to a number of themes such as highlight designing show choice optimization strategies and information pretreatment Its a put for collaboration where individuals can share their encounters inquire questions and offer feedback Analyzing the discussion strings could be exceptionally instructive in case youre fascinated by learning more almost a specific subject. [20]. Comparison of Kaggle Competition Comes about Kaggle has various competitions where diverse groups apply different machine learning and profound learning strategies to the same dataset. The comes about of these competitions give a benchmark for evaluating show execution over diverse strategies. For this examination comes about from particular dataset or competition title are compared over numerous competitions to assess the adequacy of distinctive approaches.

5 Conclusion

In terms of determining day by day and week after week commerce time arrangement think Regarding the assessment of daily and weekly commerce time series, it appears that the figure community might benefit significantly from insights derived from the Kaggle community based on our study and review of the six most recent Kaggle prediction competitions.
Our research indicates that, although time series exhibiting these characteristics are less represented within the M4 competition dataset, it does include time series that are similar to those found in the Kaggle competitions. Furthermore, the datasets from Kaggle differ from those of the M4 competition because they allow access to external information, like corporate structure or external factors, which significantly enhances forecasting precision.We find that global ensemble models yield better results than local single models, which aligns with the findings from the M4 competition. In the latest four Kaggle competitions, machine learning methods have significantly outperformed traditional time series and statistical approaches, unlike the M4 competition and the previous two Kaggle challenges. This observation can be attributed to how machine learning strategies utilize external data to cross-train and assess the influence of exogenous variables. Additionally, we note a similarity

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 2282-2285

between the two leading solutions in the M4 competition, which relied on neural networks or gradient boosted decision trees, and the top entries in the Kaggle competitions. Nevertheless, it is necessary to implement various adjustments to the machine learning methods and their validation processes to fully gain the performance advantages they offer.We strongly encourage the forecasting community to participate in the ongoing enhancement of machine learning methods for time series forecasting and to draw insights from them. Since the forecasting task and dataset share considerable similarities with certain Kaggle contests evaluated in this paper, the M5 competition serves as an excellent opportunity for this. We believe that the learnings derived from the Kaggle contests discussed in this paper will have predictive value regarding the outcomes of the M5 competition.The limitations arising from Kaggle's focus on providing solutions for practical forecasting challenges rather than academic competitions result in some constraints on the conclusions researchers can make from these contests. Without access to the test set following the end of a competition, it is unfeasible to investigate significant differences in the performance of solutions or to evaluate performance metrics diversely. Moreover, it becomes challenging to provide a comparative analysis of different solutions on varying data subsets to better recognize the pros and cons of alternative methods.Although Kaggle encourages participants to share their solutions, there is no requirement for them to publicly disclose their code or solutions, and the lack of available public solutions impacts the dependability of our evaluation. Our findings might differ if the top 25 undisclosed solutions employed techniques such as local time series models or linear regression. Given the datasets' discontinuous nature and the effects of external variables, we conclude that local time series had very slim chances of succeeding in the last four competitions. This conclusion is supported by the benchmarking results. Furthermore, we identify a systematic reporting bias resulting from inconsistencies in individual willingness to share for various reasons.

1 References

1. Al-Taie, M. Z., Salim, N., & Obasa, A. I. (2017). Successful Data Science Projects: Lessons Learned from Kaggle Competition. Kurdistan Journal of Applied Research, 2(3), 40-49.
2. Wang, A. Y., Wang, D., Drozdal, J., Liu, X., Park, S., Oney, S., & Brooks, C. (2021, May). What makes a well-documented notebook? a case study of data scientists' documentation practices in kaggle. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (pp. 1-7).
3. Banachewicz, K., & Massaron, L. (2022). The Kaggle Book: Data analysis and machine learning for competitive data science. Packt Publishing Ltd.
4. Rather, I. H., Kumar, S., & Gandomi, A. H. (2024). Breaking the data barrier: a review of deep learning techniques for democratizing AI with small datasets. Artificial Intelligence Review, 57(9), 226.
5. Owolabi, D. O., & Onobhayedo, P. (2024). A Comparative Study of Supervised Machine Learning for Effective Bots Accounts Detection on Kaggle.
6. Skiena, S. S. (2017). The data science design manual. Springer.
7. Hassan, E., Shams, M. Y., Hikal, N. A., & Elmougy, S. (2023). COVID-19 diagnosis-based deep learning approaches for COVIDx dataset: A preliminary survey. Artificial intelligence for disease diagnosis and prognosis in smart healthcare, 107-122.
8. Oala, L., Aversa, M., Nobis, G., Willis, K., Neuenschwander, Y., Buck, M., ... & Sanguinetti, B. (2022). Data models for dataset drift controls in machine learning with optical images. arXiv preprint arXiv:2211.02578.
9. AlFarsi, G., Tawafak, R. M., ElDow, A., Malik, S. I., Jabbar, J., & Al Sideiri, A. (2021). Smart classroom technology in artificial intelligence: A review paper. In International Conference on Culture Heritage, Education, Sustainable Tourism, and Innovation Technologies (pp. 229-235).
10. Al Sideiri, A., Alzeidi, N., Al Hammoshi, M., Chauhan, M. S., & AlFarsi, G. (2020). CUDA implementation of fractal image compression. Journal of Real-Time Image Processing, 17(5), 1375-1387.
11. Tawafak, R. M., Alfarsi, G., Romli, A., Jabbar, J., Malik, S. I., & Alsideiri, A. (2020, September). A Review Paper on Student-Graduate Advisory Expert system. In 2020 International Conference on Computing and Information Technology (ICCIT-1441) (pp. 1-5). IEEE.
12. Alfarsi, G., Yusof, A. B. M., Tawafak, R. M., Malik, S. I., Mathew, R., & Ashfaque, M. W. (2020, December). Instructional use of virtual reality in E-learning environments. In 2020 IEEE International Conference on Advent Trends in Multidisciplinary Research and Innovation (ICATMRI) (pp. 1-5). IEEE.
13. Alfarsi, G., Jabbar, J., Tawafak, R. M., Alsidiri, A., & Alsinani, M. (2019, December). Techniques for face verification: Literature review. In 2019 International Arab Conference on Information Technology (ACIT) (pp. 107-112). IEEE.
14. Al Farsi, G., Jabbar, J., & Tawafak, R. M. (2019, February). A Review on models of human face verification techniques. In 2019 International Conference on Fourth Industrial Revolution (ICFIR) (pp. 1-5). IEEE.
15. AlFarsi, G., & Yusof, A. B. M. (2020, November). Virtual reality applications in education domain. In 2020 21st International Arab Conference on Information Technology (ACIT) (pp. 1-7). IEEE.
16. Januschowski, T., Wang, Y., Torkkola, K., Erkkilä, T., Hasson, H., & Gasthaus, J. (2022). Forecasting with trees. International Journal of Forecasting, 38(4), 1473-1481.
17. Januschowski, T., Wang, Y., Gasthaus, J., Rangapuram, S., Türkmen, C., Zschiegner, J., ... & Schelter, S. (2024). A flexible forecasting stack. Proceedings of the VLDB Endowment, 17(12), 3883-3892.
18. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 Competition: 100,000 time series and 61 forecasting methods. International Journal of Forecasting, 36(1), 54-74.
19. Darin, S. G., & Stellwagen, E. (2020). Forecasting the M4 competition weekly data: Forecast Pro's winning approach. International Journal of Forecasting, 36(1), 135-141.
20. Al Farsi, G. (2023). The Efficiency of UTAUT2 Model in Predicting Student's Acceptance of Using Virtual Reality Technology. International Journal of Interactive Mobile Technologies, 17(12).