



Building Resilient Machine Learning Pipelines for Continuous Inference in Real-time Applications

Bhageerath Bogi

Independent Researcher, USA

DOI: <https://doi.org/10.47760/ijcsmc.2024.v13i10.008>

Abstract:

Building strong pipelines for real-time data analytics requires the convergence of AI and data engineering. In order to build scalable, effective, and robust data pipelines that enable real-time AI-driven analytics, this paper examines the architecture, technologies, and methodology required. Scalable and flexible deep learning algorithms are essential for fulfilling the ever-increasing needs of large-scale machine learning (ML) systems in the era of big datasets and real-time applications. This study examines the latest advancements in scalable deep learning techniques, paying special emphasis to architectural innovations that enable efficient model training, adaptive learning, and inference over distributed systems. As a result, it works especially well for imaging tissue slides at high resolution over a large field of view. Numerical multi-look has recently been introduced to the traditional phase retrieval procedure to make the optical setup resilient against misalignment-induced artefacts. This allows the removal of associated phase errors, albeit at the expense of a lengthy calculation time. The majority of machine learning models have a noteworthy level of resilience to input fluctuations, according to recent research on interpretable machine learning. This implies that approximation input characteristics may be accommodated by machine learning models with little to no influence on accuracy. In this research, we present Biathlon, a new machine learning serving system that finds the best degree of approximation for each aggregate characteristic by using the intrinsic robustness of models. This method guarantees a constraint on accuracy loss while allowing for maximal speedup. We test Biathlon on actual pipelines from data science contests and business applications, showing that it can exceed real-time latency constraints by attaining 5.3× to 16.6× speedup with almost no loss of accuracy.

Keywords: Machine Learning (ML), Data Pipelines, Accuracy Loss, Large-Scale, Building Robust, Real-Time Applications, Retrieval Process, Biathlon, Misalignments-Induced, High-Resolution.

I. INTRODUCTION

Both functional and temporal accuracy are essential for the design of real-time systems (RTS) [1]. Therefore, temporally predictable and deterministic algorithms are often used in the design of real-time systems. For example, [1, 2] conventional real-time scheduling algorithms must conduct (precise or adequate just) deterministic (finite time) offline feasibility (also called schedulability) tests [2, 3] prior to putting into practice an online scheduler. However, the underlying system heterogeneity and inter- and intra-dependent tasks make the scheduling algorithms' feasibility test very difficult, if not impossible, to complete [3, 4]. As a result, RTS was limited to safety- and mission-critical systems up until recently, such spacecraft and avionics, with specific proprietary hardware platforms and straightforward job models. Nevertheless, RTS has been widely used in many fields, including healthcare (implanted devices), transportation (autonomous vehicles), smart cities (smart grids), and industrial settings (drones, robots, etc.) as a result of the revolution of embedded cyber physical systems and the internet of things (IoT) (due to the quick development of hardware, software, and communication technologies) [3, 4]. These increasingly intricate cyber-physical systems have a basic resemblance in that they use a huge number of on-board sensors to interact with the physical environment with exceptional efficiency [3, 4]. In order to handle the enormous volume of varied data from on-board sensors, the systems need a lot of processing power [4]. There are a few issues with the new RTS apps:

- a) The resource-sharing models of RTS are complicated by the need for a heterogeneous hardware platform made up of CPUs (multi-core), GPUs, or specialised accelerators. [4, 5].
- b) Task dependence creates task chains with different arrival rates, which calls for complex workload models like GANG task models, DAGs, [5], and others. It is time-consuming to schedule such a task on a heterogeneous hardware platform utilising a deterministic feasibility test [5, 6].

It is impossible to overestimate the significance of data architecture in machine learning. Any successful machine learning project is built on a solid architecture that facilitates effective data management, processing, and model deployment. But there are a lot of obstacles in developing systems that can evolve with rising data volumes and model complexity [5, 6]. Organisations' approaches to data infrastructure have had to change in recent years due to the exponential rise of data. The worldwide data sphere is expected to increase from 33 zettabytes in 2018 to 175 zettabytes by 2025, according an IDC analysis [6, 7]. Traditional data infrastructures are under unprecedented strain as a result of this enormous data inflow and the growing complexity of machine learning algorithms [5, 6]. For machine learning

models to function at their best, especially deep learning architectures, enormous volumes of high-quality data are needed. As mentioned by,

"The availability of large datasets that are labelled is the most crucial element in deep learning success." [5]

But having access to big datasets alone isn't enough; companies also need to be able to effectively ingest, store, analyse, and feed this data to machine learning models in both training and production settings [6, 7].

The former, such as histology slides, is extensively used in clinical practice for the functional study of tissues [5, 6]. Staining the tissue or adding markers increases the test's specificity at the expense of its invasiveness [4]. Furthermore, light bleaching may significantly impair fluorescence's performance, and it only permits qualitative tests [5, 6]. The second class of techniques is non-specific and uses an endogenous contrast mechanism based on the refractive index that differentiates the sample and its surroundings [6, 7]. Measuring the phase delay the specimen introduces on the probing wave front not only permits morphological testing but also provides quantitative data on the specimen's optical thickness, bio volume, and dry mass [8].

In this regard, excellent examples include phase-shifting interferometry, digital holographic microscopy, spatial light interference microscopy, and Transport of Intensity Equation (TIE) imaging [8, 9]. All of the aforementioned methods have one thing in common: the optical system must be engineered to guarantee a practical trade-off between Field of View (FoV) and spatial resolution. This is especially crucial when the specimen is a tissue because imaging a large area without compromising resolution of the smallest details is essential for studying the interaction between various sample elements, investigating the non-local effects of drugs and specific treatments, and enabling robust classification of phenotypes and specimen states, such as when malignant cells are detected and metastases develop in histopathological screenings [5, 9]. In order to get Gig Pixel QPI, Fourier Psychographic Microscopy (FPM) uses a synthetic aperture approach; that is, a huge space-bandwidth product is used to extract the sample's complex amplitude [9, 10]. In order to accomplish large FoV imaging with computational super-resolution that surpasses the limitations imposed by the optical system, FPM uses low coherence light sources to probe the specimen from various angles [10].

The industry's exponential expansion in data collection and the quick development of technology have increased need for scalable and flexible deep learning algorithms [11,12]. Large-scale machine learning systems are essential in fields like natural language processing,

autonomous systems, healthcare, and finance due to their enormous dataset processing capabilities. Scalability and flexibility are crucial for ensuring efficacy and accuracy in real-world applications since the sheer volume and complexity of these datasets usually proves to be too much for conventional deep learning models [12]. As a result, researchers are focussing on developing state-of-the-art deep learning methods that can swiftly adjust to shifting computing circumstances and shifting data patterns [12, 13].

Finding a balance between model complexity and computing efficiency is a major challenge in the creation of scalable deep learning systems [13]. Although traditional deep learning architectures, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), may achieve high accuracy [14], their use in large-scale systems is constrained by their high processing power and memory needs. To solve this, a variety of techniques have been put forward, including distributed computing, model compression, and optimisation algorithms. Recent developments in parallel processing with GPUs and TPUs have accelerated the deployment of deep learning models across large-scale systems by enabling the models to handle millions or even billions of parameters [11].

1.1 Importance of Real-Time Data Analytics

Organisations can react to changes immediately thanks to real-time data analytics, which processes and analyses data as it is created with the least amount of delay [11, 12]. Real-time analytics allows for instant insights, which are crucial for many time-sensitive applications, in contrast to typical batch processing, which collects and processes data at predetermined intervals. These consist of:

- Financial services fraud detection, where real-time analysis assists in spotting fraudulent transactions as they happen [12].
- In manufacturing, predictive maintenance enables businesses to keep an eye on the operation of their equipment and take proactive measures to fix problems before they cause expensive downtime.
- Real-time analysis of consumer behaviour in e-commerce to provide customised product suggestions or promotional offers is known as personalised marketing [5].

Businesses may be proactive instead than reactive using real-time information. By empowering businesses to act on real-time data, it may lower risks, increase operational efficiency, and improve consumer experiences [11, 12]. Real-time data analytics is technically difficult to achieve, however, and calls for sophisticated analytics tools, reliable infrastructure, and effective data pipelines [11].

II. FUNDAMENTALS OF DATA ENGINEERING FOR REAL-TIME ANALYTICS

The foundation of real-time data analytics is data engineering. It entails planning and building strong, scalable, and efficient systems and procedures to manage massive data quantities [12]. In order to provide organisations relevant insights, it is important to make sure that the appropriate data is gathered, processed, and made accessible for real-time analytics. The fundamentals of data engineering will be discussed in this part, with an emphasis on data input, transformation, storage, and the crucial elements of scalability and dependability [12, 13].

2.1 Data Ingestion Technique

The act of gathering information from several sources and transferring it to a central system for further processing and analysis is known as data ingestion [13]. For real-time analytics to guarantee that data is accessible for analysis right away, this procedure has to occur with the least amount of delay possible. There are two main methods for data ingestion: [14], real-time (or streaming) processing, and batch processing.

- **Batch Processing:** Large amounts of data are gathered, and discrete batches of data are processed at predetermined times [14, 15]. This approach lacks the immediacy necessary for real-time decision-making, but it works well for analysing big datasets that don't need real-time insights.
- **Real-Time (Streaming) Processing:** With this approach, new data is instantly available as it is continually consumed as it is created. Applications where immediate insights are critical, such as fraud detection [15], predictive maintenance, and personalised marketing, need real-time ingestion [15].
- **Techniques and Tools for Real-Time Data Ingestion:** Message Brokers and Streaming Platforms: Real-time data intake is often accomplished using technologies like Rabbit MQ, Apache Kafka, and Apache Pulsar [15, 16]. The constant flow of data from many sources, including as sensors, web servers, mobile applications, and databases, is made possible by these technologies [16].
- **API Integration:** Real-time data collection from external sources is often accomplished using APIs (application programming interfaces) [17]. APIs are perfect for integrating third-party data streams, such social media feeds or financial sector data, since they allow data to move across systems seamlessly.

- **Change Data Capture (CDC):** One method for recording and monitoring database changes in real time is called CDC [17]. This enables systems to react instantly to changes in data (like an update to a user profile).

2.2 Resilient Data Ingestion Pipelines

Use an automated data lineage solution to map and comprehend the dependencies in your data flow. A complicated data pipeline has many dependencies that are impossible for a person to keep track of, and documentation, when it is there, seldom completes the task [17, 18]. Without this background, your data team will either make mistakes or be afraid to make any changes for fear of generating problems later. For further perspective, see our post on 17 data lineage use cases [18].

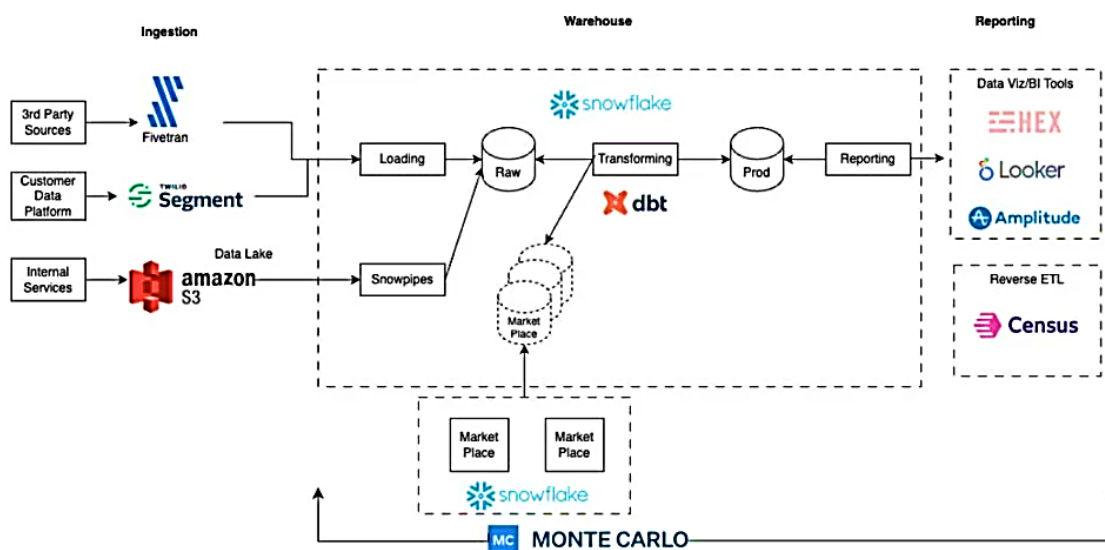


Fig. 1 Pipelines for Resilient Data Ingestion. [18]

- **Design your data pipeline so it is modular and automated:** A straightforward, modular pipeline that is easy to modify is often preferable to an ideal pipeline that must be entirely redesigned every time data changes at the source since your team will need to make modifications frequently [18, 19]. The cost of maintenance should be considered even if open source data pipeline technologies are appealing from a financial standpoint [19]. Usually, integrations lag, and teams may invest more effort in their architecture than reap the benefits.
- **Create data pipeline SLAs (service level agreements):** Aligning the design of the data pipeline with the use case is essential [19]. Do your customers need data to be updated every hour, day, minute, or second? Whether your data pipeline design uses stream, micro-batch, or batch ingestion will probably depend on that [19, 20]. Is great data quality required, or will directionally precise data be adequate? All parties are held responsible when these expectations are codified [19, 20].

- **Let the data drive the data pipeline architecture:** A data warehouse-based data pipeline design makes sense since most data teams will be working with mostly structured data for analytical reasons. A data lake may be taken into consideration by some data teams that handle more unstructured data for scientific use cases [20, 21]. Given what Netflix Studios Senior Information Engineer Dao Mi told us about his experience at Nauto, which creates AI software for car safety, others could need to design more specialised structures [22].
- **Create data products:** Zhamak Dehgani's data mesh approach [22, 23] created a new way of thinking known as "data-as-a-product." A data product is a data asset that has substantial value for the business, usually in the form of crucial tables or dashboards [24]. Discoverable, secure, controlled, reliable (high data quality), and interoperability are all characteristics of data products [23].
- **Continuously review and optimize costs:** In addition to being costly, poorly constructed and degraded queries may also lead to problems with data reliability [25]. See some useful tips for figuring out these troublesome queries and maximising Snowflake costs [25, 26].
- **Make pipelines idempotent:** You may avoid inconsistent or duplicate data by not making your pipelines idempotent, which means that performing an action more than once yields the same result [26, 27].

III. BIATHLON

Biathlon aims to run an optimised inference pipeline G with low execution cost in order to get an inference result \hat{y} that fulfils the accuracy guarantee defined in Equation 1, [27], given pipeline inputs (such as user ID), [26], an error bound δ , and a confidence level τ :

$$\Pr(|Y - \hat{y}| \leq \delta) \geq r \dots 1$$

The Biathlon procedure is shown in Figure 3. Biathlon is made up of two parts:

- (1) The Planner, which is in charge of creating an approximate online plan for the inference pipeline's execution, [27], and
- (2) The Executor, who is responsible for carrying out the inference pipeline roughly in line with the Planner's suggested plan [27, 28].

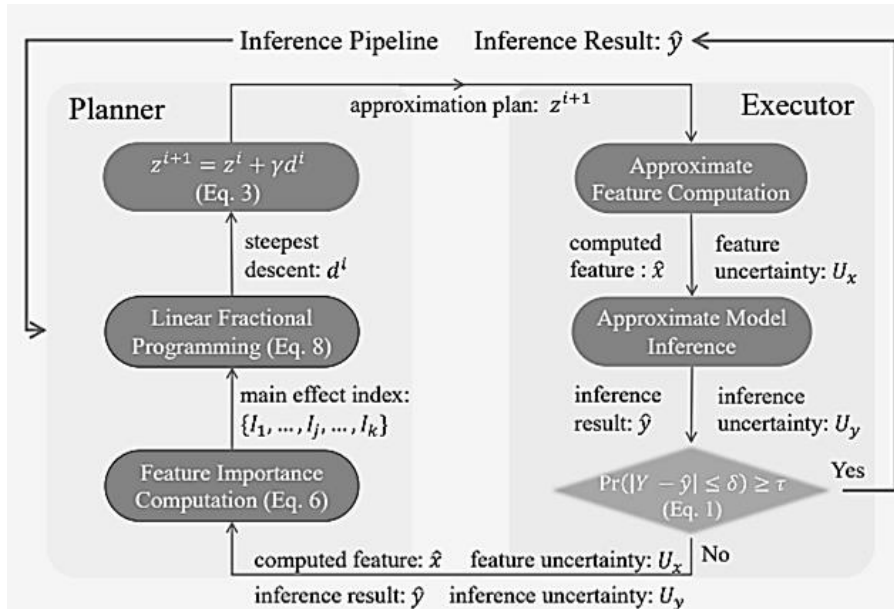


Fig. 2 System Structure of Biathlon. [27, 28]

3.1 Approximate Feature Computation (AFC)

During this phase, Biathlon determines the feature values; for non-targeting features, Biathlon determines their precise values; and for the targeting aggregation feature j , [28], Biathlon functions similarly to online aggregation, offering effective estimates for the aggregation values via a trio of steps [28].

3.2 Approximate Model Inference (AMI)

Biathlon's AMI stage has two functions:

- (1) Computing the (approximation) inference results \hat{y} using the approximation characteristics and
- (2) Estimating the uncertainty of the approximation inference result U_y [27, 28].

Biathlon computes the approximation inference result \hat{y} in a simple way: $\hat{y} = M(\hat{x})$, where M is the model inference operator, and the approximate feature values \hat{x} are immediately included into the model inference operator [22, 28].

$$d^i = \frac{\operatorname{argmax}_{\Delta z} \operatorname{Var}(Y|z^i) - \operatorname{Var}(Y|z^i + \Delta z)}{\|\Delta z\|_1} \dots\dots 2$$

$$I_j = \frac{\operatorname{Var} X_j (E - X_j(Y|X_j))}{\operatorname{Var}(Y)} \dots\dots 3$$

$$I_j^i = \frac{\operatorname{Var}(Y|z^i) - E(\operatorname{Var}(Y|z_{j*}^i))}{\operatorname{Var}(Y|z^i)} \dots\dots 4$$

Table 1 The Real –Interface Pipelines. [28]

Pipeline (description)	Data Set (Num of Records)	Num of Operators			No. of Features		No. of User Requests	
		Data store		Transformation	Model Inferences	AGG		Non-AGG
		AGG	others					
Trip-Fare (predict fare of a trip)	NYC Taxi (3B)	2	2	0	LGBM (Regression)	2	3	1541
Tick-Price (Forecast price of a tick)	Forex Tick (1.3B)	0	1	5	LR (Regression)	6	4	1696
Battery (predict remaining time to charge battery)	NASA BATTERY (7.9M)	1	0	2	LGBM (Regression)	5	8	1495
Turbofan (predict remaining useful life of turbofan)	Turbofan (54M)	9	4	0	Random Forest (Regression)	9	0	689
Bearing-Imbalance (Defect Imbalance of Bearing)	Machinery Fault (96M)	9	0	4	MLP (Classification)	9	0	326
Fraud-Detection (detect Fraudulent click of user)	Talking Data Click (325M)	9	1	0	XGB (Classifications)	3	6	1795
Student-QA (predict correctness of a questions)	Game Log (54M)	11	2	2	Random Forest (Classification)	21	2	410

IV. EVOLUTION

To show that Biathlon can reduce inference latency while maintaining acceptable accuracy loss, we evaluated it on seven real inference pipelines from Kaggle and Feast [28, 29]. Our findings indicate that using Biathlon reduces inference latency by 5.3× to 16.6× times when compared to the baseline, which is running the inferring pipeline without any approximated performance [29].

- **Workload:** Very few of the inference pipelines with costly aggregations that have been the subject of multiple papers have their matching actual data made publicly accessible [30]. The ones in FE Bench don't have any model (i.e., no trained model or training labels supplied), simply feature preparation operators [30, 31]. All of the ones we utilised for the assessment are openly accessible [31, 32].
- **System Setup:** Python is used to implement Biathlon. Our data store was Click House, [33], an open-source OLAP DBMS that supports online sampling and is intended for real-time data analytics. Biathlon is not limited to any particular data storage solution, however, and may be used with other databases, like MySQL, or data analytics frameworks, such Pandas and Dask, [33] without compromising its benefits [32, 33]. Servers with an Intel Xeon E5-2620 CPU (2.1 GHz with 8 physical cores), 256 GB of RAM, and 745 GB of Intel DC S3610 Series SSD were used for all of the studies [33].

- **Metrics:** In order to provide a thorough assessment of Biathlon, we conduct the tests five times and report the system's average latency while handling all actual user requests in addition to its speedup over the baseline [32, 34].
- **Default Configuration:** We have used a default configuration for Biathlon during an assessment process, which is shared by all workloads [33]. For the first plan, we specifically set the proportion of samples ratio at $\alpha = 0.05$ [31]. We set the step size γ to 1% of the total number of records across all features, in accordance with standard online aggregation systems. The value of the confidence level τ is 0.95 [32].

4.1 End to End Performance

Biathlon's performance assessment on the seven inference pipelines in the default setup is shown in Figure 4. The latency comparison between the baseline, RALF, and Biathlon on the seven different workloads is shown in the top figure [22, 32]. It is clear that Feature Computation (FC) takes the longest and accounts for the majority of the baseline's delay [32].

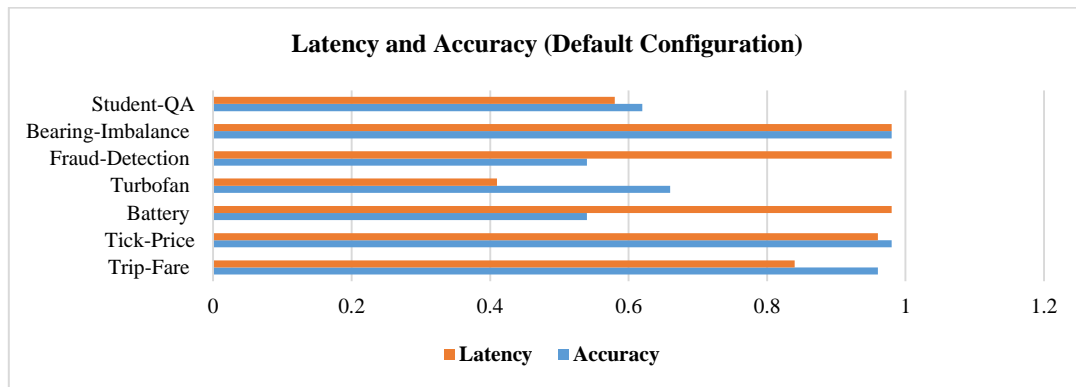


Fig. 3 Accuracy and Latency (Default Setting). [35, 36]

4.2 Varying the error bound δ

Our goal in this experiment is to investigate the effects of Equation 1's error limit δ [36, 37]. The accuracy and speedup findings in respect to different error limit values δ are shown in Figure 7. Similar to the previous example, Biathlon's accuracy is determined by utilising the precise value that the baseline anticipated as the oracle label [38, 39]. Since the others require categorisation, which cannot accept any mistake, only the output of regression pipelines is shown [35].

Biathlon is generally not advised for deep model pipelines [34, 35]. In comparison to conventional non-deep models, deep learning models are often computationally costly. The cost resulting from Biathlon's numerous model inferences for each inferences request during quasi-Monte Carlo (QMC) [34] may be more than the advantages of feature approximation.

Nonetheless, it is noteworthy that there are matching solutions available and that speeding deep learning pipelines is a significant topic of study [40].

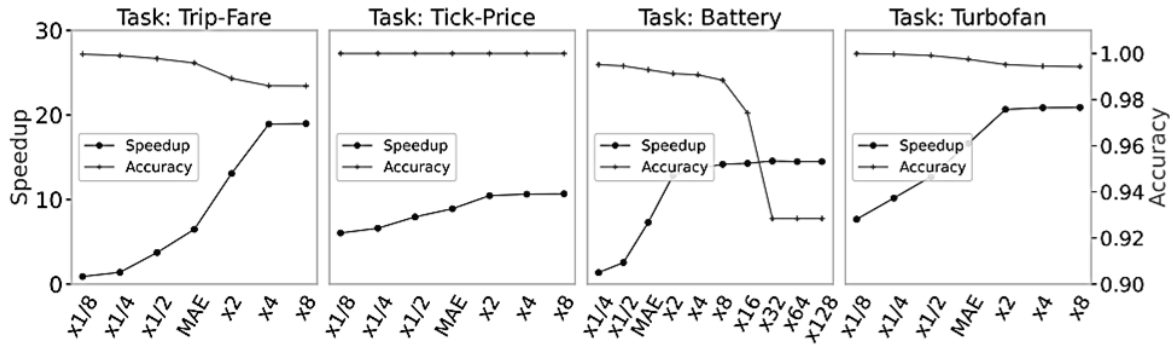


Fig. 4 Varying Error Bound δ (Regression Only). [35, 39]

V. CONCLUSION

We presented an overview of the most advanced machine learning algorithms used in embedded and real-time Internet of Things systems in this work. To sum up, in the big data age, scalable data architectures are essential to the success of machine learning projects. As data volumes increase and machine learning tasks get more complicated, scalable and adaptable deep learning algorithms become more and more important. In this study, we have looked at the advancements in flexible strategies for handling large-scale machine learning systems and scalable deep learning methods. Using appropriate methods for data collection, transformation, and validation, the system's data flow may be effectively handled. Additionally, the development of an efficient pipeline model is aided by exception and real-time testing processes as well as data quality monitoring.

In this study, we introduce Biathlon, a novel machine learning serving system designed for industrial inference pipelines and data science. Designed to meet the demanding user-facing latency requirements of real-time inference, Biathlon consists of a number of essential elements: an iterative optimisation algorithm for identifying the optimal approximation plan; uncertainty propagation from statistical analysis to estimate inference uncertainty; feature importance based on Sobol Indices from model interpretability to evaluate the contribution of individual features to the inference uncertainty; and approximate query processing from the database area to compute feature approximately.

Biathlon delivers a speedup ranging from $5.3\times$ to $16.6\times$ on actual pipelines without a discernible decrease in accuracy, and it gives the greatest speedup with a probabilistic assurance of limited error. Some operators, such Top-K, are inherited from online

aggregation and are not approximated by Biathlon. We think that the feature store method, as RALF has shown, may be a good substitute in certain situations. But it's important to note that the feature store approach—which includes RALF—has drawbacks as well, including not having error boundaries or being constrained to a certain range of workloads.

REFERENCES

- [1]. Burns and R. Davis, "Mixed criticality systems-a review," Department of Computer Science, University of York, Tech. Rep, pp. 1–69, 2013.
- [2]. D. Michie, D. J. Spiegelhalter, and C. C. Taylor, "Machine learning, neural and statistical classification," 1994.
- [3]. S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 science and information conference. IEEE, 2014, pp. 372–378.
- [4]. M. S. Mahdavinejad, M. Rezvan, M. Barekatin, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine learning for internet of things data analysis: A survey," Digital Communications and Networks, vol. 4, no. 3, pp. 161–175, 2018.
- [5]. M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," Nature Reviews Genetics, vol. 16, no. 6, pp. 321–332, 2015.
- [6]. N. Sebe, I. Cohen, A. Garg, and T. S. Huang, Machine learning in computer vision. Springer Science & Business Media, 2005, vol. 29.
- [7]. R. Caruana and A. Niculescu-Mizil, "An empirical comparison of supervised learning algorithms," in Proceedings of the 23rd international conference on Machine learning, 2006, pp. 161–168.
- [8]. T. Hastie, R. Tibshirani, and J. Friedman, "Overview of supervised learning," in The elements of statistical learning. Springer, 2009, pp. 9–41.
- [9]. H. B. Barlow, "Unsupervised learning," Neural computation, vol. 1, no. 3, pp. 295–311, 1989.
- [10]. T. Hastie, R. Tibshirani, and J. Friedman, "Unsupervised learning," in The elements of statistical learning. Springer, 2009, pp. 485–585.
- [11]. L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," Journal of artificial intelligence research, vol. 4, pp. 237–285, 1996.
- [12]. T. Radil, P. M. Ramos, F. M. Janeiro, and A. C. Serra, "Pq monitoring system for real-time detection and classification of disturbances in a single-phase power system," IEEE Transactions on Instrumentation and Measurement, vol. 57, no. 8, pp. 1725–1733, 2008.
- [13]. S. Vadineanu and M. Nasri, "Robust and accurate period inference ~ using regression-based techniques," in IEEE Real-Time Systems Symposium (RTSS), 2020, pp. 358–370.
- [14]. B. Brandenburg, J. M. Calandrino, and J. H. Anderson, "On the scalability of real-time scheduling algorithms on multicore platforms: A case study," in 2008 Real-Time Systems Symposium. IEEE, 2008, pp. 157–169.
- [15]. A. Predescu, C. Negru, M. Mocanu, and C. Lupu, "Real-time clustering for priority evaluation in a water distribution system," in 2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR). IEEE, 2018, pp. 1–6.
- [16]. Valencia and A. Alimohammad, "A real-time spike sorting system using parallel osort clustering," IEEE transactions on biomedical circuits and systems, vol. 13, no. 6, pp. 1700–1713, 2019.
- [17]. V. Bianco et al., "Strategies for reducing speckle noise in digital holography," Light: Sci. Appl., vol. 7, pp. 1–16, 2018.
- [18]. S. Montrésor et al., "Comparative study of multi-look processing for phase map de-noising in digital fresnel holographic interferometry," JOSA A, vol. 36, pp. A59–A66, 2019.
- [19]. V. Bianco et al., "Self-propelling bacteria mimic coherent light decorrelation," Opt. Exp., vol. 23, pp. 9388–9396, 2015.
- [20]. V. Bianco et al., "Random resampling masks: A non-Bayesian one-shot strategy for noise reduction in digital holography," Opt. Lett., vol. 38, pp. 619–621, 2013.
- [21]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778.
- [22]. W. Yang et al., "Deep learning for single image super-resolution: A brief review," IEEE Trans. Multimedia, vol. 21, no. 12, pp. 3106–3121, May 2019.
- [23]. L. Chen et al., "Deep learning in label-free cell classification," Sci. Rep., vol. 6, pp. 1–16, 2016.
- [24]. Y. Li et al., "Deep cytometry: Deep learning with real-time inference in cell sorting and flow cytometry," Sci. Rep., vol. 9, pp. 1–12, 2019.
- [25]. C. Tian et al., "Deep learning on image denoising: An overview," Neural Netw., vol. 131, pp. 251–75, 2020.
- [26]. T. Nguyen, Y. Xue, Y. Li, L. Tian, and G. Nehmetallah, "Deep learning approach for Fourier psychographic microscopy," Opt. Exp., vol. 26, 2018, Art. no. 26470.
- [27]. J. Zhang, T. Xu, Z. Shen, Y. Qiao, and Y. Zhang, "Fourier psychographic microscopy reconstruction with multiscale deep residual network," Opt. Exp., vol. 27, pp. 8612–8625, 2019.
- [28]. M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014.
- [29]. C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 4681–4690.

- [30].J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3155–3164.
- [31].Patel, M., & Kumar, R. (2016). Data engineering frameworks for big data analytics. *International Journal of Data Science and Analytics*, 2(1), 43-56.
- [32].Wang, J., & Zhao, L. (2018). Integrating AI with data engineering: Challenges and opportunities. *Data & Knowledge Engineering*, 113, 1-12.
- [33].Aghera, S. (2011). Design and Development of Video Acquisition System for Aerial. *Management*, 41(4), 605-615.
- [34].Aghera, S. (2011). Design and development of video acquisition system for aerial surveys of marine animals. Florida Atlantic University.
- [35].Kalva, H., Marques, O., Aghera, S., Reza, W., Giusti, R., & Rahman, A. Design and Development of a System for Aerial Video Survey of Large Marine Animals. Muthu, P., Mettikolla, P., Calander, N., Luchowski, R., Gryczynski, I., Gryczynski, Z., & Borejdo, J. (2010). Single molecule kinetics in the familial hypertrophic cardiomyopathy D166V mutant mouse heart. *Journal of molecular and cellular cardiology*, 48(5), 989-998.
- [36].Krupa, A., Fudala, R., Stankowska, D., Loyd, T., Allen, T. C., Matthay, M. A., & Kurdowska, A. K. (2009). Anti-chemokine autoantibody: chemokine immune complexes activate endothelial cells via IgG receptors. *American journal of respiratory cell and molecular biology*, 41(2), 155-169.
- [37].Joseph M. Hellerstein, Peter J. Haas, and Helen J. Wang. 1997. Online Aggregation. In Proceedings of the 1997 ACM SIGMOD International Conference on Management of Data (SIGMOD '97). Association for Computing Machinery, New York, NY, USA, 171–182.
- [38].Zezhou Huang, Rathijit Sen, Jiaxiang Liu, and Eugene Wu. 2023. JoinBoost: Grow Trees Over Normalized Data Using Only SQL. (Jul 2023).
- [39].JS. 2023. Forex tick data huge database since April 2020.
- [40].Daniel Kang, John Emmons, Firas Abuzaid, Peter Bailis, and Matei Zaharia. 2017. NoScope: Optimizing Neural Network Queries over Video at Scale. *Proc. VLDB Endow.* 10, 11 (aug 2017), 1586–1597.
- [41].Naveen Bagam, International Journal of Computer Science and Mobile Computing, Vol.13 Issue.11, November- 2024, pg. 6-27
- [42].Naveen Bagam. (2024). Optimization of Data Engineering Processes Using AI. *International Journal of Research Radicals in Multidisciplinary Fields*, ISSN: 2960-043X, 3(1), 20–34. Retrieved from <https://www.researchradicals.com/index.php/rr/article/view/138>
- [43].Naveen Bagam. (2024). Machine Learning Models for Customer Segmentation in Telecom. *Journal of Sustainable Solutions*, 1(4), 101–115. <https://doi.org/10.36676/j.sust.sol.v1.i4.42>
- [44].Bagam, N. (2023). Implementing Scalable Data Architecture for Financial Institutions. *Stallion Journal for Multidisciplinary Associated Research Studies*, 2(3), 27
- [45].Bagam, N. (2021). Advanced Techniques in Predictive Analytics for Financial Services. *Integrated Journal for Research in Arts and Humanities*, 1(1), 117–126. <https://doi.org/10.55544/ijrah.1.1.16>
- [46].Enhancing Data Pipeline Efficiency in Large-Scale Data Engineering Projects. (2019). *International Journal of Open Publication and Exploration*, ISSN: 3006-2853, 7(2), 44-57. <https://ijope.com/index.php/home/article/view/166>
- [47].Sai Krishna Shiramshetty. (2024). Enhancing SQL Performance for Real-Time Business Intelligence Applications. *International Journal of Multidisciplinary Innovation and Research Methodology*, ISSN: 2960-2068, 3(3), 282–297. Retrieved from <https://ijmirm.com/index.php/ijmirm/article/view/138>
- [48].Sai Krishna Shiramshetty, "Big Data Analytics in Civil Engineering : Use Cases and Techniques", International Journal of Scientific Research in Civil Engineering (IJSRCE), ISSN : 2456-6667, Volume 3, Issue 1, pp.39-46, January-February.2019
URL : <https://ijsrce.com/IJSRCE19318>
- [49].Sai Krishna Shiramshetty, " Data Integration Techniques for Cross-Platform Analytics, International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCEIT), ISSN : 2456-3307, Volume 6, Issue 4, pp.593-599, July-August-2020. Available at doi : <https://doi.org/10.32628/CSEIT2064139>
- [50].Shiramshetty, S. K. (2021). SQL BI Optimization Strategies in Finance and Banking. *Integrated Journal for Research in Arts and Humanities*, 1(1), 106–116. <https://doi.org/10.55544/ijrah.1.1.15>
- [51].Sai Krishna Shiramshetty. (2022). Predictive Analytics Using SQL for Operations Management. *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal*, 11(2), 433–448. Retrieved from <https://eduzonejournal.com/index.php/eiprmj/article/view/693>
- [52].Shiramshetty, S. K. (2023). Data warehousing solutions for business intelligence. *International Journal of Computer Science and Mobile Computing*, 12(3), 49–62. <https://ijcsmc.com/index.php/volume-12-issue-3-march-2023/>
- [53].Sai Krishna Shiramshetty. (2024). Comparative Study of BI Tools for Real-Time Analytics. *International Journal of Research and Review Techniques*, 3(3), 1–13. Retrieved from <https://ijrtr.com/index.php/ijrtr/article/view/210>
- [54].Sai Krishna Shiramshetty "Leveraging BI Development for Decision-Making in Large Enterprises" *Iconic Research And Engineering Journals Volume 8 Issue 5 2024 Page 548-560*
- [55]. Sai Krishna Shiramshetty "Integrating SQL with Machine Learning for Predictive Insights" *Iconic Research And Engineering Journals Volume 1 Issue 10 2018 Page 287-292*
- [56].Shiramshetty, S. K. (2023). Advanced SQL Query Techniques for Data Analysis in Healthcare. *Journal for Research in Applied Sciences and Biotechnology*, 2(4), 248–258. <https://doi.org/10.55544/jrasb.2.4.33>
- [57].Naveen Bagam. (2024). Data Integration Across Platforms: A Comprehensive Analysis of Techniques, Challenges, and Future Directions. *International Journal of Intelligent Systems and Applications in Engineering*, 12(23s), 902–919. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/7062>

- [58].Naveen Bagam, Sai Krishna Shiramshetty, Mouna Mothey, Harish Goud Kola, Sri Nikhil Annam, & Santhosh Bussa. (2024). Advancements in Quality Assurance and Testing in Data Analytics. *Journal of Computational Analysis and Applications (JoCAAA)*, 33(08), 860–878. Retrieved from <https://eudoxuspress.com/index.php/pub/article/view/1487>
- [59].Bagam, N., Shiramshetty, S. K., Mothey, M., Kola, H. G., Annam, S. N., & Bussa, S. (2024). Optimizing SQL for BI in diverse engineering fields. *International Journal of Communication Networks and Information Security*, 16(5). <https://ijcnis.org/>
- [60].Bagam, N., Shiramshetty, S. K., Mothey, M., Annam, S. N., & Bussa, S. (2024). Machine Learning Applications in Telecom and Banking. *Integrated Journal for Research in Arts and Humanities*, 4(6), 57–69. <https://doi.org/10.55544/ijrah.4.6.8>
- [61].Bagam, N., Shiramshetty, S. K., Mothey, M., Kola, H. G., Annam, S. N., & Bussa, S. (2024). Collaborative approaches in data engineering and analytics. *International Journal of Communication Networks and Information Security*, 16(5). <https://ijcnis.org/>
- [62].Kola, H. G. (2024). Optimizing ETL Processes for Big Data Applications. *International Journal of Engineering and Management Research*, 14(5), 99–112. <https://doi.org/10.5281/zenodo.14184235>
- [63].SQL in Data Engineering: Techniques for Large Datasets. (2023). *International Journal of Open Publication and Exploration, ISSN: 3006-2853*, 11(2), 36-51. <https://ijope.com/index.php/home/article/view/165>
- [64].Data Integration Strategies in Cloud-Based ETL Systems. (2023). *International Journal of Transcontinental Discoveries, ISSN: 3006-628X*, 10(1), 48-62. <https://internationaljournals.org/index.php/ijtd/article/view/116>
- [65].Harish Goud Kola. (2024). Real-Time Data Engineering in the Financial Sector. *International Journal of Multidisciplinary Innovation and Research Methodology, ISSN: 2960-2068*, 3(3), 382–396. Retrieved from <https://ijmirm.com/index.php/ijmirm/article/view/143>
- [66].Harish Goud Kola. (2022). Best Practices for Data Transformation in Healthcare ETL. *Edu Journal of International Affairs and Research, ISSN: 2583-9993*, 1(1), 57–73. Retrieved from <https://edupublications.com/index.php/ejiar/article/view/106>
- [67].Kola, H. G. (2018). Data warehousing solutions for scalable ETL pipelines. *International Journal of Scientific Research in Science, Engineering and Technology*, 4(8), 762. <https://doi.org/10.1.1.123.4567>
- [68].Harish Goud Kola, " Building Robust ETL Systems for Data Analytics in Telecom , International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 5, Issue 3, pp.694-700, May-June-2019. Available at doi : <https://doi.org/10.32628/CSEIT1952292>
- [69].Kola, H. G. (2022). Data security in ETL processes for financial applications. *International Journal of Enhanced Research in Science, Technology & Engineering*, 11(9), 55. <https://ijsrcseit.com/CSEIT1952292>.
- [70].Annam, S. N. (2020). Innovation in IT project management for banking systems. *International Journal of Enhanced Research in Science, Technology & Engineering*, 9(10), 19. https://www.erpublications.com/uploaded_files/download/sri-nikhil-annam_gBNPz.pdf
- [71].Annam, S. N. (2018). Emerging trends in IT management for large corporations. *International Journal of Scientific Research in Science, Engineering and Technology*, 4(8), 770. <https://ijsrset.com/paper/12213.pdf>
- [72].Sri Nikhil Annam, " IT Leadership Strategies for High-Performance Teams, International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 7, Issue 1, pp.302-317, January-February-2021. Available at doi : <https://doi.org/10.32628/CSEIT228127>
- [73].Annam, S. N. (2024). Comparative Analysis of IT Management Tools in Healthcare. *Stallion Journal for Multidisciplinary Associated Research Studies*, 3(5), 72–86. <https://doi.org/10.55544/sjmars.3.5.9>.
- [74].Annam, N. (2024). AI-Driven Solutions for IT Resource Management. *International Journal of Engineering and Management Research*, 14(6), 15–30. <https://doi.org/10.31033/ijemr.14.6.15-30>
- [75].Annam, S. N. (2022). Optimizing IT Infrastructure for Business Continuity. *Stallion Journal for Multidisciplinary Associated Research Studies*, 1(5), 31–42. <https://doi.org/10.55544/sjmars.1.5.7>
- [76].Sri Nikhil Annam , " Managing IT Operations in a Remote Work Environment, International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 8, Issue 5, pp.353-368, September-October-2022. <https://ijsrcseit.com/paper/CSEIT23902179.pdf>
- [77].Annam, S. (2023). Data security protocols in telecommunication systems. *International Journal for Innovative Engineering and Management Research*, 8(10), 88–106. <https://www.ijiemr.org/downloads/paper/Volume-8/data-security-protocols-in-telecommunication-systems>
- [78].Annam, S. N. (2023). Enhancing IT support for enterprise-scale applications. *International Journal of Enhanced Research in Science, Technology & Engineering*, 12(3), 205. https://www.erpublications.com/uploaded_files/download/sri-nikhil-annam_urfNc.pdf
- [79].Santhosh Bussa, "Advancements in Automated ETL Testing for Financial Applications", *IJRAR - International Journal of Research and Analytical Reviews (IJRAR)*, E-ISSN 2348-1269, P- ISSN 2349-5138, Volume.7, Issue 4, Page No pp.426-443, November 2020, Available at : <http://www.ijrar.org/IJRAR2AA1744.pdf>
- [80].Bussa, S. (2023). Artificial Intelligence in Quality Assurance for Software Systems. *Stallion Journal for Multidisciplinary Associated Research Studies*, 2(2), 15–26. <https://doi.org/10.55544/sjmars.2.2.2>.
- [81].Bussa, S. (2021). Challenges and solutions in optimizing data pipelines. *International Journal for Innovative Engineering and Management Research*, 10(12), 325–341. <https://sjmars.com/index.php/sjmars/article/view/116>
- [82].Bussa, S. (2022). Machine Learning in Predictive Quality Assurance. *Stallion Journal for Multidisciplinary Associated Research Studies*, 1(6), 54–66. <https://doi.org/10.55544/sjmars.1.6.8>
- [83].Bussa, S. (2022). Emerging trends in QA testing for AI-driven software. *International Journal of All Research Education and Scientific Methods (IJARESM)*, 10(11), 1712. Retrieved from <http://www.ijaesm.com>
- [84].Santhosh Bussa. (2024). Evolution of Data Engineering in Modern Software Development. *Journal of Sustainable Solutions*, 1(4), 116–130. <https://doi.org/10.36676/j.sust.sol.v1.i4.43>

- [85].Santhosh Bussa. (2024). Big Data Analytics in Financial Systems Testing. *International Journal of Multidisciplinary Innovation and Research Methodology*, ISSN: 2960-2068, 3(3), 506–521. Retrieved from <https://ijmirm.com/index.php/ijmirm/article/view/150>
- [86].Bussa, S. (2019). AI-driven test automation frameworks. *International Journal for Innovative Engineering and Management Research*, 8(10), 68–87. Retrieved from <https://www.ijemr.org/public/uploads/paper/427801732865437.pdf>
- [87].Santhosh Bussa. (2023). Role of Data Science in Improving Software Reliability and Performance. *Edu Journal of International Affairs and Research*, ISSN: 2583-9993, 2(4), 95–111. Retrieved from <https://edupublications.com/index.php/ejar/article/view/111>
- [88].Bussa, S. (2023). Enhancing BI tools for improved data visualization and insights. *International Journal of Computer Science and Mobile Computing*, 12(2), 70–92. <https://doi.org/10.47760/ijcsmc.2023.v12i02.005>.
- [89].Sai Krishna Shiramshetty. (2024). Enhancing SQL Performance for Real-Time Business Intelligence Applications. *International Journal of Multidisciplinary Innovation and Research Methodology*, ISSN: 2960-2068, 3(3), 282–297. Retrieved from <https://ijmirm.com/index.php/ijmirm/article/view/138>
- [90].Sai Krishna Shiramshetty, "Big Data Analytics in Civil Engineering : Use Cases and Techniques", International Journal of Scientific Research in Civil Engineering (IJSRCE), ISSN : 2456-6667, Volume 3, Issue 1, pp.39-46, January-February.2019
URL : <https://ijsrce.com/IJSRCE19318>
- [91].Sai Krishna Shiramshetty, " Data Integration Techniques for Cross-Platform Analytics, International Journal of Scientific Research in Computer Science, Engineering and Information Technology(IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 4, pp.593-599, July-August-2020. Available at doi : <https://doi.org/10.32628/CSEIT2064139>
- [92].Shiramshetty, S. K. (2021). SQL BI Optimization Strategies in Finance and Banking. *Integrated Journal for Research in Arts and Humanities*, 1(1), 106–116. <https://doi.org/10.55544/ijrah.1.1.15>
- [93].Sai Krishna Shiramshetty. (2022). Predictive Analytics Using SQL for Operations Management. *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal*, 11(2), 433–448. Retrieved from <https://eduzonejournal.com/index.php/eiprmj/article/view/693>
- [94].Sai Krishna Shiramshetty. (2024). Comparative Study of BI Tools for Real-Time Analytics. *International Journal of Research and Review Techniques*, 3(3), 1–13. Retrieved from <https://ijrrt.com/index.php/ijrrt/article/view/210>
- [95]. Sai Krishna Shiramshetty "Leveraging BI Development for Decision-Making in Large Enterprises" Iconic Research And Engineering Journals Volume 8 Issue 5 2024 Page 548-560
- [96].Shiramshetty, S. K. (2023). Advanced SQL Query Techniques for Data Analysis in Healthcare. *Journal for Research in Iconic Research And Engineering Journals Applied Sciences and Biotechnology*, 2(4), 248–258. <https://doi.org/10.55544/jrasb.2.4.33>
- [97].Sai Krishna Shiramshetty "Integrating SQL with Machine Learning for Predictive Insights" Iconic Research And Engineering Journals Volume 1 Issue 10 2018 Page 287-292
- [98].Shiramshetty, S. K. (2023). Advanced SQL Query Techniques for Data Analysis in Healthcare. *Journal for Research in Applied Sciences and Biotechnology*, 2(4), 248–258. <https://doi.org/10.55544/jrasb.2.4.33>
- [99].Shiramshetty, S. K. (2023). Advanced SQL Query Techniques for Data Analysis in Healthcare. *Journal for Research in Applied Sciences and Biotechnology*, 2(4), 248–258. <https://doi.org/10.55544/jrasb.2.4.33>
- [100]. Mouna Mothey. (2022). Automation in Quality Assurance: Tools and Techniques for Modern IT. *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal*, 11(1), 346–364. Retrieved from <https://eduzonejournal.com/index.php/eiprmj/article/view/694>
- [101]. Mothey, M. (2022). Leveraging Digital Science for Improved QA Methodologies. *Stallion Journal for Multidisciplinary Associated Research Studies*, 1(6), 35–53. <https://doi.org/10.55544/sjmars.1.6.7>
- [102]. Mothey, M. (2023). Artificial Intelligence in Automated Testing Environments. *Stallion Journal for Multidisciplinary Associated Research Studies*, 2(4), 41–54. <https://doi.org/10.55544/sjmars.2.4.5>
- [103]. Mouna Mothey. (2024). Test Automation Frameworks for Data-Driven Applications. *International Journal of Multidisciplinary Innovation and Research Methodology*, ISSN: 2960-2068, 3(3), 361–381. Retrieved from <https://ijmirm.com/index.php/ijmirm/article/view/14>
- [104]. Bagam, N., Shiramshetty, S. K., Mothey, M., Annam, S. N., & Bussa, S. (2024). Machine Learning Applications in Telecom and Banking. *Integrated Journal for Research in Arts and Humanities*, 4(6), 57–69. <https://doi.org/10.55544/ijrah.4.6.8>
- [105]. Harish Goud Kola. (2024). Real-Time Data Engineering in the Financial Sector. *International Journal of Multidisciplinary Innovation and Research Methodology*, ISSN: 2960-2068, 3(3), 382–396. Retrieved from <https://ijmirm.com/index.php/ijmirm/article/view/143>
- [106]. Harish Goud Kola. (2022). Best Practices for Data Transformation in Healthcare ETL. *Edu Journal of International Affairs and Research*, ISSN: 2583-9993, 1(1), 57–73. Retrieved from <https://edupublications.com/index.php/ejar/article/view/106>