

Cloud-Based Data Analytics Models for Scalable Computing Systems

Prof. Peter Sloot

Professor, University of Amsterdam, Netherlands

Abstract

The rapid growth of digital information has created substantial demand for computing infrastructures capable of storing, processing, and analyzing large and continuously generated datasets. Conventional computing environments may experience limitations when workloads increase rapidly or require real-time analytical processing. Cloud computing provides an elastic and distributed infrastructure through which organizations can dynamically access computing, storage, networking, and analytical resources. This study examines cloud-based data analytics models for scalable computing systems and develops a conceptual framework connecting cloud infrastructure, distributed data processing, analytical models, resource management, scalability, and intelligent decision-making. A conceptual and descriptive research design is adopted through a structured synthesis of secondary scholarly literature concerning cloud computing, Big Data Analytics, distributed computing, cloud-native architectures, machine learning, stream processing, serverless computing, and scalable data management. Existing research indicates that cloud environments can provide elasticity, flexible resource allocation, and scalable processing capabilities for large-scale analytics workloads. Cloud-supported Big Data Analytics has been increasingly associated with distributed processing frameworks, real-time analytics, data-intensive applications, and machine-learning workloads. The study analyzes major cloud-based analytics models, including Infrastructure-as-a-Service, Platform-as-a-Service, serverless computing, cloud-native distributed processing, and hybrid cloud architectures.

The analysis identifies scalability, processing performance, cost efficiency, fault tolerance, data security, privacy, interoperability, resource utilization, and workload variability as major factors influencing model effectiveness. The findings suggest that scalable cloud analytics requires an integrated architecture in which data storage, processing frameworks, analytical algorithms, and resource management mechanisms operate cohesively. Recent research further indicates that cloud-native and distributed frameworks can address growing data volumes through parallel processing, elastic resource allocation, containerization, and event-driven architectures. The study concludes that cloud-based data analytics can provide a flexible foundation for scalable computing systems when organizations appropriately match architectural models with workload characteristics, analytical requirements, security obligations, and cost constraints.

Keywords: Cloud Computing, Data Analytics, Scalable Computing, Big Data Analytics, Cloud-Native Architecture, Distributed Computing, Serverless Computing, Machine Learning, Data Processing, Elastic Computing

1. Introduction

The rapid expansion of digital technologies has resulted in unprecedented growth in the volume, velocity, and variety of data generated by organizations, connected devices, digital platforms, enterprise applications, social networks, and scientific systems. Traditional computing infrastructures may struggle to process such information efficiently because they are often designed around fixed computational capacity. When workloads increase significantly, organizations may need to purchase additional hardware, expand storage systems, and manually configure computing resources. Cloud computing has emerged as an important alternative by providing on-demand access to distributed computing and storage resources. Cloud environments can dynamically allocate resources according to workload requirements, creating opportunities for more flexible and scalable data analytics. Recent research identifies cloud computing as an important infrastructure for handling large-scale data analytics and real-time processing requirements.

Cloud-based data analytics refers to the use of cloud infrastructure, platforms, and services to collect, store, process, analyze, and interpret large datasets. The model can combine cloud storage with distributed processing frameworks, databases, machine-learning systems, visualization platforms, and real-time analytics tools. Instead of maintaining all computational resources internally, organizations can access computing capacity as required. This approach can be particularly useful for workloads that fluctuate significantly over time. During periods of high demand, additional resources can be provisioned, while unused capacity can be released when workloads decline. Such elasticity represents one of the fundamental advantages of cloud computing for scalable analytics.

The increasing integration of Big Data Analytics with cloud platforms has further expanded the role of cloud-based computing systems. Distributed frameworks can divide large analytical tasks across multiple processing nodes, enabling parallel computation and reducing the time required to analyze extensive datasets. Frameworks such as Hadoop and Spark have played important roles in distributed data processing, while newer cloud-native architectures increasingly incorporate containers, event-driven systems, stream processing, and serverless computing. Recent research describes the transition from traditional Hadoop-based systems toward cloud-native architectures that emphasize distributed processing, containerization, orchestration, and elastic resource management.

Despite these advantages, cloud-based analytics introduces several technical, financial, and organizational challenges. Data security and privacy are particularly important because organizations may transfer sensitive information to external cloud environments. Performance can also be affected by network latency, data movement, resource contention, and inefficient workload configuration. Cost management represents another important issue because pay-as-you-go models can become expensive when large-scale analytical jobs are executed continuously. Cloud-based analytics therefore requires careful architectural planning to balance scalability, performance, security, and cost. Recent research identifies security, privacy, resource availability, timeliness, and cost management as continuing challenges in cloud-based Big Data Analytics.

This study examines cloud-based data analytics models for scalable computing systems and develops a conceptual framework connecting cloud infrastructure with analytical performance. The study evaluates different cloud service and deployment models and examines their contributions to scalability, distributed processing, resource efficiency, real-time analytics, and intelligent decision-

making. It also analyzes the major challenges associated with cloud-based analytics and identifies the factors that influence successful implementation. The overall purpose is to establish a structured understanding of how cloud-based analytical architectures can support increasingly large, complex, and dynamic computing workloads.

2. Objectives of the Study

2.1 Primary Objective

The primary objective of this study is to examine cloud-based data analytics models that support scalable computing systems by enabling elastic resource allocation, distributed data processing, efficient storage, advanced analytics, real-time computation, intelligent decision-making, and flexible management of rapidly changing computational workloads.

2.2 Evaluation of Cloud-Based Analytics Models

The study evaluates major cloud-based analytical approaches, including Infrastructure-as-a-Service, Platform-as-a-Service, serverless computing, cloud-native distributed processing, and hybrid cloud architectures. It examines their respective capabilities in handling large datasets, variable workloads, real-time processing, machine-learning applications, and computationally intensive analytics while considering scalability, flexibility, performance, and operational requirements.

2.3 Assessment of Scalability and Performance Factors

The study identifies the principal technological and organizational factors influencing scalable cloud analytics. Particular attention is given to distributed processing, elastic resource allocation, data locality, processing latency, fault tolerance, cost management, security, privacy, interoperability, and workload variability. The objective is to establish a comprehensive conceptual perspective for designing efficient and reliable cloud-based data analytics systems.

3. Review of Literature

3.1 Cloud Computing and Big Data Analytics

Cloud computing and Big Data Analytics have developed as complementary technological domains because both address the requirements of modern data-intensive applications. Big Data environments require substantial storage and computational resources, while cloud computing provides scalable infrastructure capable of supporting such requirements. Research on Big Data Analytics in cloud computing emphasizes that the increasing volume of data has made traditional processing systems insufficient for many modern analytical workloads.

Cloud infrastructure can provide computing resources through service-oriented models that reduce the requirement for organizations to maintain extensive physical infrastructure. Infrastructure-as-a-Service allows users to access virtualized computing, storage, and networking resources. Platform-as-a-Service provides higher-level development and analytical environments, while software and serverless services abstract infrastructure management further. These approaches allow organizations to select an appropriate level of infrastructure control according to analytical requirements.

The combination of cloud infrastructure with distributed analytics frameworks can improve the processing of large datasets. Parallel computing allows workloads to be divided into smaller tasks and executed across multiple nodes. This can improve processing speed and enable organizations to analyze datasets that would be difficult to process on a single machine. Research on scalable data analysis emphasizes the importance of parallel implementations and cloud resources for large-scale and real-time data analysis.

3.2 Distributed and Cloud-Native Analytics

Distributed data processing represents an important foundation for scalable cloud analytics. Frameworks such as Hadoop, Spark, and Flink enable large datasets to be processed across multiple computing nodes. Hadoop is strongly associated with distributed batch processing, while Spark provides in-memory processing capabilities and broader support for iterative analytics. Flink is particularly relevant to stream-processing workloads that require low latency. Recent comparative research identifies scalability, processing speed, fault tolerance, resource utilization, and integration with machine learning as important characteristics of distributed data-processing frameworks.

Cloud-native architectures extend distributed computing through containerization, orchestration, microservices, event-driven processing, and automated resource management. These architectures allow analytical components to be independently deployed and scaled according to demand. This is particularly valuable for applications in which different parts of an analytical pipeline experience different workload patterns.

Serverless computing represents another development in cloud-native analytics. In serverless environments, organizations can execute analytical functions without directly managing underlying servers. Resources can be provisioned automatically according to requests or events. Recent research examining cloud computing for Big Data Analytics identifies serverless architectures as an increasingly important approach for scalable and cost-efficient data-intensive applications.

3.3 Scalable Analytics and Intelligent Computing

Cloud-based analytics increasingly incorporates machine learning and artificial intelligence to generate predictive and prescriptive insights. Machine-learning workloads may require substantial computational resources, particularly when models are trained on large datasets. Cloud infrastructure can provide scalable computing capacity for training and deployment while enabling organizations to access specialized resources when required.

Real-time analytics represents another important application area. Streaming data generated by sensors, financial systems, digital platforms, and IoT devices may need to be analyzed immediately rather than stored for later processing. Cloud-based streaming architectures can combine distributed message-processing systems with scalable analytical engines to support continuous data analysis. Research on modern distributed processing emphasizes the importance of stream processing for real-time and event-driven applications.

The literature also indicates that cloud-based analytics must balance scalability with security, privacy, and cost. Although cloud environments can reduce the need for peak-provisioned infrastructure, extensive workloads may still generate significant costs. Data movement between storage and

computing services can also influence performance and expenditure. Therefore, scalable cloud analytics should be designed around workload characteristics rather than assuming that unlimited cloud resources automatically produce optimal performance.

4. Research Methodology

4.1 Research Design

This study adopts a conceptual and descriptive research design to examine cloud-based data analytics models for scalable computing systems. The research is based on a structured synthesis of secondary scholarly literature concerning cloud computing, Big Data Analytics, distributed processing, cloud-native architectures, serverless computing, machine learning, stream processing, and scalable data management. The conceptual design enables the study to integrate technological and analytical perspectives into a unified framework.

4.2 Data Sources and Analytical Procedure

Relevant scholarly literature was examined according to major themes including cloud service models, distributed data processing, scalability, elasticity, real-time analytics, machine learning, resource optimization, security, privacy, and cost management. Studies were compared according to the capabilities and limitations of different cloud-based analytical approaches. Particular attention was given to research examining scalable data processing, cloud-native architectures, distributed frameworks, and large-scale analytical workloads. The identified themes were synthesized to develop a structured understanding of cloud analytics performance.

4.3 Conceptual Framework Development

The proposed framework consists of seven interconnected stages: data acquisition, cloud-based storage, distributed processing, analytical modeling, resource orchestration, performance evaluation, and decision support. Data are first collected from enterprise applications, sensors, digital platforms, and external sources. The data are stored in scalable cloud environments and processed through distributed computing frameworks. Statistical, machine-learning, and artificial-intelligence models transform processed information into analytical outputs. Cloud orchestration mechanisms dynamically allocate computational resources according to workload requirements. Performance is continuously evaluated using measures such as processing time, throughput, resource utilization, scalability, cost, and reliability.

5. Analysis

5.1 Cloud Service Models and Analytical Scalability

Cloud-based data analytics can be implemented through different service models, each providing a different level of infrastructure management. Infrastructure-as-a-Service provides organizations with greater control over virtual machines, storage, networking, and computing configurations. This approach can be suitable for organizations that require customized analytical environments or need control over distributed processing frameworks. However, it also requires greater responsibility for infrastructure configuration, maintenance, and security.

Platform-as-a-Service provides a higher level of abstraction by offering managed development and analytical environments. Organizations can focus more directly on analytical applications rather than managing underlying infrastructure. This can reduce operational complexity and support faster deployment of data-processing and machine-learning systems.

Function-as-a-Service or serverless computing provides an even greater degree of abstraction. Computational functions are executed in response to events or requests, and infrastructure resources are automatically managed by the cloud provider. Recent empirical research comparing cloud service models for Big Data Analytics reports that serverless configurations can be particularly suitable for modular and burst-oriented workloads, while infrastructure-based models can remain useful for continuous workloads.

5.2 Distributed Processing and Resource Optimization

Distributed processing enables cloud systems to divide large analytical workloads among multiple computational nodes. This approach can significantly improve scalability because additional nodes can be introduced when processing requirements increase. Distributed architectures also provide opportunities for parallel computation, allowing independent tasks to execute simultaneously.

Resource optimization is critical because cloud environments operate according to dynamically allocated resources. Poorly configured workloads may consume excessive computational capacity, increasing costs without producing proportional performance benefits. Efficient resource allocation can involve workload scheduling, autoscaling, data partitioning, caching, load balancing, and computational resource matching.

Cloud-native systems can further improve resource utilization through containerization and orchestration. Containers allow analytical components to operate in isolated environments, while orchestration platforms can automatically manage deployment and scaling. These capabilities are particularly useful for complex analytics pipelines containing multiple independent processing stages.

Table 1: Comparison of Cloud-Based Data Analytics Models

Cloud Analytics Model	Scalability	Infrastructure Control	Major Advantage	Major Limitation
IaaS	High	High	Customization	Higher management responsibility
PaaS	High	Moderate	Simplified development	Platform dependency
SaaS Analytics	High	Low	Easy accessibility	Limited customization
Serverless/FaaS	Very High for burst workloads	Low	Automatic scaling	Cold-start and execution constraints

Cloud-Native Containers	Very High	Moderate-High	Flexible deployment	Orchestration complexity
Hybrid Cloud	High	High	Security and flexibility balance	Integration complexity
Multi-Cloud	Very High	High	Provider diversification	Management complexity

5.3 Real-Time Analytics and Intelligent Computing

Real-time analytics requires systems capable of processing information continuously as it arrives. Cloud-based stream-processing architectures can support such workloads by distributing incoming events across processing nodes. Applications include fraud detection, industrial monitoring, recommendation systems, traffic management, cybersecurity, and IoT analytics. Real-time systems must minimize latency while maintaining reliability and processing accuracy.

Machine learning can be integrated into cloud analytics pipelines to generate predictive insights from both historical and streaming information. Models can be trained using large-scale cloud resources and deployed through scalable inference services. This architecture can support applications where analytical models need to process large numbers of requests simultaneously.

The integration of cloud analytics with artificial intelligence can also support intelligent resource management. Analytical systems can monitor workload characteristics and predict future resource requirements. Automated resource-management mechanisms can then adjust computational capacity according to expected demand. This creates an adaptive architecture in which analytics and infrastructure management continuously influence one another.

Table 2: Key Performance Dimensions of Scalable Cloud Analytics

Performance Dimension	Measurement Focus	Importance
Processing Speed	Execution time	Faster analytical results
Throughput	Data processed per unit time	Large-scale workload handling
Scalability	Performance as resources increase	Supports workload growth
Elasticity	Speed of resource adjustment	Handles changing demand
Resource Utilization	CPU, memory, and storage usage	Improves efficiency
Cost Efficiency	Cost per analytical workload	Controls expenditure
Fault Tolerance	Recovery from failures	Improves reliability
Latency	Delay in data processing	Critical for real-time analytics
Security	Protection of data and infrastructure	Reduces risk

6. Discussion

6.1 Scalability and Operational Efficiency

- Cloud platforms allow computational resources to expand according to workload requirements.
- Distributed processing enables large datasets to be divided across multiple computing nodes.
- Elastic infrastructure reduces the need for maintaining permanent peak capacity.
- Autoscaling can respond to changes in workload intensity.
- Cloud storage can accommodate rapidly increasing datasets.
- Parallel processing can reduce execution time for large analytical workloads.
- Resource monitoring can identify inefficient computational usage.
- Cloud-based analytics can support organizations with fluctuating data-processing requirements.

6.2 Integration with Artificial Intelligence and Big Data

- Cloud platforms provide computational resources for large-scale machine learning.
- Big Data frameworks enable distributed processing of heterogeneous datasets.
- AI models can be deployed through scalable cloud services.
- Stream processing can support real-time intelligent analytics.
- Cloud storage enables centralized access to large analytical datasets.
- Distributed computing can accelerate model training.
- Cloud-native architectures can support modular AI pipelines.
- Automated resource allocation can improve AI workload efficiency.

6.3 Strategic Implications for Scalable Computing

The findings indicate that cloud-based data analytics should be designed around workload characteristics rather than a single universal architecture. Batch workloads may benefit from distributed processing frameworks optimized for large datasets, while real-time applications require low-latency stream-processing architectures. Burst-oriented workloads may benefit from serverless computing, whereas continuous workloads may require persistent infrastructure. Therefore, architecture selection should consider workload duration, data volume, latency requirements, security constraints, computational intensity, and cost.

Scalability should also be evaluated as a multidimensional concept. A system that can technically process more data is not necessarily scalable if costs increase disproportionately or performance deteriorates under heavy workloads. Effective scalability should maintain acceptable performance while resource requirements increase. Consequently, organizations should evaluate processing speed, throughput, resource utilization, cost, fault tolerance, and latency together.

Another important implication concerns hybrid and multi-cloud architectures. Organizations with strict data governance requirements may retain sensitive workloads in private infrastructure while using public cloud resources for scalable analytics. Hybrid architectures can therefore provide a balance between control and elasticity. However, integration between different environments introduces additional complexity related to networking, identity management, data movement, monitoring, and application portability.

7. Challenges

One of the major challenges associated with cloud-based data analytics is maintaining predictable performance under rapidly changing workloads. Although cloud environments provide elasticity, analytical performance can still be influenced by network latency, resource contention, data movement, storage architecture, and workload configuration. Large datasets may require substantial movement between storage and processing services, increasing processing time and costs. Organizations therefore need appropriate data partitioning, caching, processing locality, and workload scheduling strategies to maintain efficient performance.

Security and privacy represent another significant challenge. Cloud-based analytics may involve sensitive organizational, financial, healthcare, customer, or operational information. Moving such data to shared or externally managed infrastructures creates concerns about unauthorized access, data breaches, identity management, encryption, and regulatory compliance. Organizations must implement strong authentication, access controls, encryption, monitoring, and governance mechanisms. Privacy-preserving analytical techniques may also become important when multiple organizations share data across cloud environments.

Cost management is a further challenge. The pay-as-you-go model can provide economic advantages when resources are used efficiently, but uncontrolled workloads can generate substantial expenses. Continuous data processing, large storage volumes, frequent data transfers, and computationally intensive machine-learning workloads can increase expenditure. Organizations therefore need cost-monitoring mechanisms, resource optimization strategies, workload scheduling, and appropriate service selection. Cloud analytics should be evaluated according to total analytical value rather than infrastructure flexibility alone.

Another challenge concerns interoperability and technological complexity. Cloud analytics environments frequently combine databases, storage systems, distributed processing frameworks, machine-learning tools, APIs, containers, orchestration systems, and external services. Integrating these components can require specialized technical skills and increase system-management complexity. Vendor dependence may also create difficulties when organizations attempt to move workloads between providers. Recent research on scalable cloud-native processing emphasizes the continuing importance of flexible distributed architectures, interoperability, and efficient resource management.

8. Limitations

This study has several limitations. First, the research adopts a conceptual and descriptive methodology based on secondary scholarly literature rather than original experimental measurements. Therefore, the comparisons between cloud-based analytics models represent conceptual assessments rather than universal performance rankings. Actual performance can vary according to cloud provider, workload type, dataset characteristics, network conditions, application architecture, and configuration.

A second limitation concerns the rapid development of cloud technologies. Cloud platforms, serverless architectures, container orchestration systems, distributed processing frameworks, and artificial-intelligence services continue to evolve. New services and architectures may change the relative advantages of different analytical models. Consequently, the proposed framework should be viewed as adaptable rather than permanently fixed.

Another limitation is that cloud scalability involves multiple dimensions, including computational, storage, network, financial, and organizational scalability. A system may demonstrate strong computational scalability while experiencing limitations in network performance or cost efficiency. The present study provides a broad framework but does not empirically measure every dimension under controlled experimental conditions.

Finally, the study does not focus on a specific industry. Requirements for healthcare analytics, financial systems, manufacturing, e-commerce, scientific computing, and smart-city applications can differ substantially. Security, latency, compliance, data volume, and availability requirements may therefore influence architecture selection differently across sectors. Future research should conduct industry-specific empirical studies to validate the proposed framework.

9. Conclusion

Cloud-based data analytics has emerged as an important technological approach for supporting scalable computing systems in environments characterized by rapidly increasing data volumes and changing computational requirements. Cloud infrastructure provides organizations with access to distributed storage and computing resources, while elastic architectures enable computational capacity to adjust according to workload requirements. These characteristics make cloud platforms particularly relevant for large-scale analytics, machine learning, real-time processing, and data-intensive applications.

The analysis demonstrates that different cloud-based analytics models provide different combinations of scalability, control, flexibility, cost, and management responsibility. Infrastructure-as-a-Service provides substantial control, Platform-as-a-Service simplifies analytical development, serverless architectures support highly elastic workloads, and cloud-native distributed systems provide flexible processing capabilities. Hybrid cloud models can further balance scalability with security and governance requirements. Consequently, architecture selection should be based on workload characteristics and organizational priorities.

Scalable cloud analytics also requires effective distributed processing and resource management. Frameworks such as Hadoop, Spark, and Flink demonstrate different strengths for batch, iterative, and streaming workloads, while modern cloud-native architectures extend scalability through containers, orchestration, event-driven systems, and automated resource allocation. Recent research confirms the continuing evolution toward distributed, cloud-native, and increasingly energy-efficient architectures for large-scale analytics.

Overall, cloud-based data analytics can provide a strong foundation for scalable, flexible, and intelligent computing systems when technological architecture is appropriately aligned with workload characteristics, security requirements, performance objectives, and cost constraints. Future research should focus on adaptive resource management, sustainable cloud computing, privacy-preserving analytics, multi-cloud interoperability, intelligent workload scheduling, and AI-driven cloud optimization. Such developments can further improve the ability of cloud environments to support increasingly complex and data-intensive computing applications.

10. Declaration

The author declares that this research was prepared independently for academic and scholarly purposes and that no external organization influenced the research design, interpretation, or conclusions. The study is based on secondary scholarly literature and does not involve the collection of confidential personal information or direct experimentation with human participants. No financial or commercial relationship influenced the preparation or presentation of this manuscript. The research follows principles of academic integrity, transparency, responsible data management, and objective scholarly interpretation. The author accepts responsibility for the accuracy and academic presentation of the information contained in this manuscript.

References

1. Ng, A. Y. (2004). Feature selection, L1 vs. L2 regularization, and rotational invariance. *Proceedings of the 21st International Conference on Machine Learning*.
2. Belkin, M., & Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6), 1373–1396.
3. Hanneke, S. (2016). The optimal sample complexity of PAC learning. *Journal of Machine Learning Research*, 17, 1–15.
4. Kaufmann, E., Cappé, O., & Garivier, A. (2016). On the complexity of best-arm identification in multi-armed bandit models. *Journal of Machine Learning Research*, 17, 1–42.
5. Khaleghi, A., Ryabko, D., Mary, J., & Preux, P. (2016). Consistent algorithms for clustering time series. *Journal of Machine Learning Research*, 17, 1–32.
6. Blaser, R., & Fryzlewicz, P. (2016). Random rotation ensembles. *Journal of Machine Learning Research*, 17, 1–26.
7. Mentch, L., & Hooker, G. (2016). Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *Journal of Machine Learning Research*, 17, 1–41.
8. Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., & Schölkopf, B. (2016). Distinguishing cause from effect using observational data: Methods and benchmarks. *Journal of Machine Learning Research*, 17, 1–102.
9. Meng, X., Bradley, J., Yavuz, B., et al. (2016). MLlib: Machine learning in Apache Spark. *Journal of Machine Learning Research*, 17, 1–7.
10. Kpotufe, S., Boularias, A., Schultz, T., & Kim, K. (2016). Gradients weights improve regression and classification. *Journal of Machine Learning Research*, 17, 1–34.
11. Martínez, A. M., Webb, G. I., Chen, S., & Zaidi, N. A. (2016). Scalable learning of Bayesian network classifiers. *Journal of Machine Learning Research*, 17, 1–35.
12. Doğan, Ü., Glasmachers, T., & Igel, C. (2016). A unified view on multi-class support vector classification. *Journal of Machine Learning Research*, 17, 1–32.
13. Lu, J., Hoi, S. C. H., Wang, J., Zhao, P., & Liu, Z.-Y. (2016). Large scale online kernel learning. *Journal of Machine Learning Research*, 17, 1–43.
14. Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18, 1–5.

15. Kucukelbir, A., Tran, D., Ranganath, R., Gelman, A., & Blei, D. M. (2017). Automatic differentiation variational inference. *Journal of Machine Learning Research*, 18, 1–45.
16. Wainer, J., & Cawley, G. (2017). Empirical evaluation of resampling procedures for optimising SVM hyperparameters. *Journal of Machine Learning Research*, 18, 1–35.
17. Ito, N., Takeda, A., & Toh, K.-C. (2017). A unified formulation and fast accelerated proximal gradient method for classification. *Journal of Machine Learning Research*, 18, 1–49.
18. Vehtari, A., Mononen, T., Tolvanen, V., Sivula, T., & Winther, O. (2016). Bayesian leave-one-out cross-validation approximations for Gaussian latent variable models. *Journal of Machine Learning Research*, 17, 1–38.
19. Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, 17, 1–40.
20. Russo, D., & Van Roy, B. (2016). An information-theoretic analysis of Thompson sampling. *Journal of Machine Learning Research*, 17, 1–30.
21. Sutton, R. S., Mahmood, A. R., & White, M. (2016). An emphatic approach to the problem of off-policy temporal-difference learning. *Journal of Machine Learning Research*, 17, 1–29.
22. Damianou, A. C., Titsias, M. K., & Lawrence, N. D. (2016). Variational inference for latent variables and uncertain inputs in Gaussian processes. *Journal of Machine Learning Research*, 17, 1–62.
23. Zuluaga, M., Krause, A., & Püschel, M. (2016). e-PAL: An active learning approach to the multi-objective optimization problem. *Journal of Machine Learning Research*, 17, 1–32.
24. Babbar, R., Partalas, I., Gaussier, E., Amini, M.-R., & Amblard, C. (2016). Learning taxonomy adaptation in large-scale classification. *Journal of Machine Learning Research*, 17, 1–37.
25. Richtárik, P., & Takáč, M. (2016). Distributed coordinate descent method for learning with big data. *Journal of Machine Learning Research*, 17, 1–25.