Journal of Discrete Mathematical Sciences & Cryptography
ISSN 0972-0529 (Print), ISSN 2169-0065 (Online)
Vol. 26 (2023), No. 3, pp. 641–655
DOI : 10.47974/JDMSC-1735

Quantum-aided feature selection model – A quantum machine learning approach

Rupam Bhagawati *
Thiruselvan Subramanian †
Department of Computer Science and Engineering
Presidency University Bangalore
Bangaluru 560064
Karnataka
India

Abstract

The accuracy of information retrieval systems is measured by the relevancy of retrieved results as per the user's query. Relevant results are presented by performing various methods viz. indexing and crawling and the output of these processes is the retrieved results that have to pass through the ranking process which is the central goal of information retrieval systems. The ranking is carried out through the classification or clustering of processed results which can include redundant and noisy features. The accuracy of classification or clusters for the ranking process can be maximized by removing noisy and duplicate features through the feature selection method. Although feature selection is an expensive computational process, after many decades, quantum computation tools are in use for many algorithms to implement realistic problems, particularly in the standard of Quantum Annealing.

This paper focuses to prospect the standard of quantum computing in order to increase the quality of information classes through the feature selection method. The persuasiveness of the quantum approach is comparable to the classical process that focused on the reliability of quantum methodology from different perspectives.

Subject Classification: *81-08 : Primary 20G42, Secondary: 62J05..*

Keywords: *Information retrieval, Quantum computing, Feature selection, Quantum annealing .*

* *E-mail:* **rupambhagawati6@gmail.com (Corresponding Author)**
† *E-mail:* **thirulic@gmail.com**

© Taru Publications

1. Introduction

Information retrieval is a process with the unique responsibility of fetching relevant information to users, as per their needs effectively and efficiently. The ranking is a centric goal while performing this responsibility [1]. To achieve ranking in a proper way numerous algorithms are proposed through machine learning approaches and classification approaches as well as clustering approaches [2]. Almost all the approaches suffered from feature selection problems. As per the given query for relevant information, it is necessary to identify the relevant feature of that information corresponding to that query [3]. In today's scenario, the richness and complexity of data are increasing exponentially which requires an exponential calculation of the relevant features of data [4]. Deep learning techniques and neural approaches mitigate the problem to a great extent and presented better overall effectiveness. In spite of effectiveness, performance is not stable for the approaches and varies a lot across different topics [5]. As pointed out by Bender [6] Neural approaches are demanding on the account of computing resources and a larger amount of data which require a larger model to accomplish and are not free from risk.

Feature selection is a technique to select a subset of features from a data set that has more relevant information than the total set of data due to many reasons which include collinearity, redundancy, and noise as if constant data [7]. The process of feature selection reduced the dimension of the data set as a preprocessing step. Dimensional reduction of data set is a benefit for learning algorithms as it performs better scaling for learning algorithms and is cost-effective for training, and testing modeling [8]. Unlike feature engineering based on principal component analysis came up with higher visibility and explainability [9], feature selection can pull output from the analyzer model during the selection process described in [10] as wrapper methods. In [11] and [12] feature selection is broadly described for various situations and techniques to be implemented according to the situation.

In this paper, we think ourselves, about whether it is possible to make feature selection more accessible by using different learning algorithms as well as investigate the feasibility of its working with quantum computation. In this study, we explore checking improved performance through quantum-aided methods to find a combination of features of a subset of size k. In our study, we projected the setup by framing a binary quadratic model and a quantum device is projected to search optimal

feature subsets. To carry out the process, various distance and correlation theorems are deployed to formulate the quadratic optimization model. The maximal relevancy and minimum redundancy idea as pointed out in [13] is applied with a basis of well-being divining towards the target variable (Maximal relevancy) and less feature correlation between an independent variable (Minimum redundancy). Finally, to set a benchmark for the quantum-aided model, two regression models are selected to compare the outputs and discussed the further usability of the defined model.

2. Associated Models and Works

The feature selection process is a verse of mathematical formulation with an objective to differentiate a subset of features from a set of data. This process improves the effectiveness of a model used for various purposes viz. ranking, classification, clustering, and many more [14]. This process is effective to reduce the cost of computation and it's used widely for research [15, 16]. As per [17] in regard to classification applications requirement of feature selection is fulfilled by applying mutual information-based metrics for supervised learning. Feature selection procedure in the recent era is also carried out by boosting process for the effective ranking model in a classical methodology of the classifier [18]. Boosting procedure is explained in [19] for the mitigation of different feature selection problems during ranking and classification. Timeline of research bring quantum computation for feature selection; quantum computing is obtaining acceptance in recent years for the deployment of various algorithm in its realm. In the ideal of quantum computing, Quantum Annealing is also one of the most widely used mechanisms for various problems starting from the QUBO problems [20, 21]. QUBO formulation for different important problems has been developed viz. graph partitioning [22], support vector machine [23], and optimization of resource allocation [24]. In the verse of machine learning, recently a new correlation metric is introduced namely Maximal Information Coefficient [25] with more effective results of measuring relationships which can be incorporated with QUBO formulation. Recently, a QUBO solver is applied to avoid resonance by designing an optimal structure [26]. Working on QUBO algorithms for various classical and quantum devices achieved a successful result. In classical methodology, there is a problem of transforming of the original optimization problem to a QUBO problem which is solved by the proposal of the Black Box Optimization (BBO) method in [27].

Quantum Annealing (QA) was applied recently for collaborative filtering [28]. The availability of Quantum Processing units (QPU) online, for various quantum formulations makes it simple to develop multiple applications. Entirely QPUs are used to solve real-life problems; proved by optimizing the traffic flow in China [29] which uses the quantum annealing (QA) system.

Quantum annealing is in trend to solve feature selection problems for various situations. In [30] mutual information is used to evaluate Q matrix, which is considered the initial designing phase for the feature selection problems using a quadratic program. Q matrix is used to present quadratic terms for minimizing the multivariate objective function. It is a positive semi-definite matrix [31]. Many specific tasks put effort into feature subset selection by applying quantum annealing using mutual information, quadratic binary optimization, and linear regression [32, 33].

It is stated in [34] that for industry-relevant problems like traffic flow optimization quantum annealing processes are based on models. Quantum annealing processes for real-world problems are specifically model-oriented [35]. Quantum annealing approach for scheduling applications, graphical networking, and timeliness which are closer to real-world problems faced a limitation of available hardware size [36]. Similarly, as per [37] real Financial problems is beyond the quantum annealing due to the time-consuming behavior of the approach.

In our presented model design, to the best of our knowledge, we put effort to use various existing techniques to perform feature subset selection from the set of documents that can implement ranking in an efficient way for the information needs of people in the realm of information retrieval.

3. Proposed Feature Selection Methodology

The process of feature selection is carried out for a group or set of information through a supervised learning technique that evolves various learning algorithms. The desired goal is to reduce the conceptual errors for the given function between the information of one group and make an error-free new group of information. As to test and form a new group of data there are various learning algorithms and on the basis of different algorithms, here, an assumption is proposed that correlates with existing learning algorithms. To carry out the development of the algorithm, the dataset is represented in the form:

$$\mathcal{D} = \{X \in R^{N\,X\,M}, Y \in R^{N}\}$$

Representing $\mathcal{D}$ data matrix, X of size N rows and M columns and Y of N rows and 1 column belongs to random matrices $R^{N\,X\,M}$ and a random column matrix R^N respectively. Both X and Y matrices are row-column vectors and column vectors respectively. In the data matrix, various learning algorithms are applied to make an approximate assumption of $h(X) \simeq Y$.

For the feature selection process, initially, we choose to find the subset of the feature as X_k where each row vector $x_i \{i = 1, 2, 3, \ldots . n\}$ in X reduced column size to k states that each row is filtered down as $\{x_{i1}, x_{i2}, x_{i3}, \ldots . x_{ik}\}$ and $k < M$. To choose an optimal subset of feature or column of size k, various k values are applied to function $h(X_k) \simeq y$ in such a way that loss is lesser than or equal to $h(X) \simeq y$ and cross-verified with a different test set within the data where $h(X) \simeq Y$ is the approximate hypothesis for various learning algorithms. Performing with this problem, there is an assumption of the existence of one optimal feature subset of size k, identical for different datasets and acceptable set $h(X_k) = h(X)$ i.e., $k = M$.

Comparison among the features and target variable is carried out by examining the relationship among the features set of data matrix X. Initially, distance functions are interpreted to present the binary optimization problem for feature subset selection. The task is carried out by comparing different distance functions to represent linear and quadratic terms for the binary quadratic model.

In this study, Maximal Information Coefficient (MIC) [38] and Mutual Information [39] are investigated. As to computing features from different classes of data variables, both purely continuous as well as continuous and discrete variables are considered, to estimate probabilities for the MI calculations as there are different strategies for accumulated continuous values. To find probabilities among test and train sets, MI is defined as:

$$MI(x, y) = \Sigma_{x,y} P(x, y) \ln \frac{P(x,y)}{P(x)P(y)} \tag{i}$$

Inclusion of density estimation, continuous to discrete variables accumulation, and clustering algorithms. To estimate MI, *K-nn binning* strategy [40] is used:

$$I(x, y) = \psi(N) - \psi(x) + \psi(k) - \psi(m) \tag{ii}$$

Where, $I(x, y)$ *tends to* $MI(x, y)$, ψ is the gamma function, k is the distance between assigned neighbors, and m is the total count of neighbors [41].

Mutual Information Coefficient is also proposed as this algorithm is capable to depict mutual information to a grid and to handle purely continuous as well as continuous and discrete variables. As per [42] MIC is defined, while a characteristic matrix *C* is defined to compute MIC*:*

$$C(x,y)_{i,j} = \frac{I^*(x,y)_{i,j}}{\log_2 \min(i,j)} \tag{iii}$$

In this equation, $I^*(x,y)$ represent accumulated values of *x, y* in a grid G_{ij} (grid representation of dataset X_{ij}). So MIC using the characteristic matrix will be:

$$MIC = max_{ij<B(n)}\{C(x,y)_{i,j}\} \tag{iv}$$

In which $B(n)$ set the maximum grid size as per [25] to $n^{0.6}$. and *i, j* is a product to compile the grid size.

3.1 *Formulae QUBO*

Quadratic Binary Optimization Model (QUBO) design with a formula incorporated for the proposed method of feature selection [43]. At first, for each column vector *i, in X* contrasted to *y,* the distance measure is computed to determine the absolute value for the distance measure. Secondly, to discover a minimum for the optimization problem: the transpose of absolute value is calculated. These computed values are represented diagonally as a linear term in the Q matrix.

With these linear terms, till this point three components are equipped to describe the proposed algorithm:

1. Evaluated the Absolute value of the distance function, which is the significant weight for the features that have higher relevancy.
2. Minimum of the domain, referred by the transpose of absolute values which correspond as target variables; to maintain the correlation equally between positive and negative.
3. Lastly the linear terms diagonally through the Q matrix, are used as the key to present the portion of maximal relevancy for the model.

Now, for the minimum redundancy characteristic of the model, the pairwise distance between each column vector indexed as *i and j,* in X is evaluated. Distance values are used to formulate a relation between quadratic terms and binary quadratic models, and these values are represented on the upper diagonal part of the Q matrix.

$$Q_{ij} = \begin{cases} -|\text{distance}\ (X_i, y)|, & \text{if } i = j \\ |\text{distance}\ (X_i, X_j)|, & \text{if } i < j \\ 0, & \text{otherwise} \end{cases} \quad \text{(v)}$$

Where, (X_i, X_j) is the column vector of the document matrix (X_{ij}) and y is the set of test and train data.

QUBO formulae are obtained by determining combinations of features that are not correlated as well as have more distance from each other, that retain the relevancy to the target variables.

Steps to pursue annealing schedule and sample solutions to evaluate minimum energy (E) :

1. For annealing: Represent a vector of qubits value σ, encompasses a scaling parameter φ to balance the domain of optimization:

$$E(\sigma) = \varphi \Sigma_{i \le j} \sigma_i Q_{ij} \sigma_j \quad \sigma \in \{0,1\} \quad \text{(vi)}$$

2. Enforced restriction on the resulting sample of objective function: Maintain single value for *k qubits* = 1 by introducing a balancing parameter ϕ.

$$E(\sigma) = \varphi \Sigma_{i \le j} \sigma_i Q_{ij} \sigma_j + \phi(\Sigma_i \sigma_i - k)^2 \quad \text{(vii)}$$

Representing the resulting vector σ, of length M and a limitation that assures *k qubits* with only single value of 1 and others as 0. Finally, filtering the data set X to acquire X_k, in which k columns are processed by index location where $\sigma_m = 1$ for all $\{\sigma_1, \sigma_2, \sigma_3, ..\sigma_m\}$.

3.2 *Performance Evaluations*

Model design with discussed formulae and procedure is evaluated for its performance, initially from the representation of testing and training data. The dataset used to check the performance is arranged as a 70/30 percentage split between training and test sets. Two existing models viz. multiple linear regression and gradient boosted regression tree are employed as analyzers to investigate the working of the proposed model for feature selection.

The linear regression model is presented as a multiple linear regression model to clarify the presence of two or more independent variables to determine an absolute path to depict a target variable. As per [44] multiple linear regression model is used to predict the group of

contradictory variables against the target variables. In this study it is used to find y which entitled variable for trained data:

$$y = Xw + \varepsilon \tag{viii}$$

Where ε is the random error term and we assume that matrix X has full rank i.e. rank of X is p and w is $p \times 1$ vector of parameters. In this, w is optimized so that $\varepsilon = y - Xw$ is minimized. So the new test data X, is substituted with the trained parameter y.

Gradient boosted regression (GBR) tree [45], is deployed to analyze the model and to approximate the benchmark in contrast to the Multiple Linear regression model (LR). For both models, validation is carried out through two layers based on performance criteria for the trained model specified by Mean Absolute Error (MAE).

Representation of performance meridians:

Performance criteria are interpreted through a calculation of *hit grade* and *length score* by considering the cardinality of intersection between two sets of *index-selected features* (K_{sf}) *and optimal features* (K_{of}). The stated performance criteria determine the *FS-Accuracy*.

$$hit\ grade = \frac{cardinality(K_{Sf} \cap K_{Of})}{cardinality\ K_{Of}} \tag{ix}$$

$$length\ score = \frac{|\,cardinality(K_{Sf}) - cardinality(K_{Of})\,|}{cardinality(K_{Of})} \tag{x}$$

$$FS accuaracy = \frac{hitgrade + length\ score}{2} \tag{xi}$$

Two layers to verify the validations:

1. Complete feature selections *(CF):* Analyzer adjusted for all features of training set of X and y to initiate a baseline for all performance meridians and to recognize all test errors.
2. Feature elimination(FE): To set a standard of quantum aided method over RFE wrapper method [41] a test is performed.

3.3 *Datasets and Implementation details*

Tested the model with a dataset of documents, Reuters21578 [46] is a multi-class, multi-label dataset with 90 classes, 7769 sets of training documents, and 3019 sets of testing documents providing a total of 21578

collections. Datasets have binary accuracy and subset accuracy which is used to set the train and test set. With a vocabulary size of 35247 in the training set, the dataset is ready to use for python libraries. Even after hard decomposition or restriction on vocabulary, 1207 words remain for the training set. While using the dataset for our proposed model, we have selected all 3019 documents. The finalized set consists of different features of documents like numbers of words, numbers of stop words, and acquaintances size. After applying one hot encoding through python libraries for categorical and normal features it preserved the attribute size as 28. Table I shows the sample of the dataset used in the study.

Table I

Sample Categories of documents with training and testing values in Reuters datasets.

Category	Documents	Training	Testing
CCAT	381327	10416	23077
GCAT	239267	2050	5180
MCAT	204820	5154	11110
C15	151785	3122	7454
ECAT	119920	3162	7539
M14	85440	1799	4887
C151	81890	515	698
C152	73092	1088	435
GPOL	56878	1627	3756
M13	53634	1095	2613

Implementation is carried out by using python libraries viz. scikit-learn [47], pandas [48], and Q Feature Selection [49]. For Quantum annealing, access to the D-wave annealing machine is obtained from Leap software and API along with the diamond library.

4. Results and Discussion

Discussion towards the result starts initially from the adjustment of parameters declared in different steps. Values for σ and ϕ in the QUBO formulae will initiate the process, manually in the model we depicted the value for $\sigma = 1000$ and $\phi = 10$ which will generate an acceptable result for formulated baseline methods. The parameter σ, with mentioned value, constructs a scaling effect that will design a strong objective function to

optimize. The value of $\phi = 10$ is responsible to compose the result set from lower to higher scale by maintaining the size of k.

For the proposed model, performance meridians are investigated on the basis of Mean Absolute Error (MAE) and feature selection subset size k across test data of the dataset. The best 10 samples are considered from bootstrapped samples for quantum-aided function. Cross-validation for both cases was held to specify the results.

In both cases, the quantum-aided function and RFE function runtime for the feature selection process is considered identically. For the quantum function, anneal time was taken into consideration in Quantum Processing Unit (QPU) while for the RFE function, normal clock time was in the Central Processing Unit (CPU). Results showed that quantum annealing time remains constant with respect to problem size whereas RFE function time varied with respect to the size of input and wrapper function. Wrapper function evaluation is measured on the basis of normal clock time.

As per Table II, the result achieved by the Maximal Information coefficient (MIC) in the realm of quantum, projected in QUBO formulae is better than that of Mutual Information (MI) performance within QUBO formulae with a lowered mean absolute error (MAE) of 2.42 and accuracy (FS accuracy) of 0.9 by keeping an optimal measure of $k = 5$ for both the criteria. Finally, for the Reuters dataset we can observe the MAE= 3.12 and 3.22 for Complete feature (ALL) using LR and GBR respectively keeping $k = 28$ in the realm of quantum is lesser than MAE= 3.96 and 3.98 achieved by Feature Elimination (FE) using the same LR and GBR with the traditional RFE function keeping $k = 12$. As per the observed result, it can be stated that matrices are effective for special type dataset for text classification. Quantum-aided model for future selection methods can be suggested for

Table II

Result for Reuters dataset included MAE and Accuracy

Regression Methods	Correlation Metrics		
	MAE	k	FS Accuracy
QMI-LR	2.48 ± 0.12	5	0.82
QMIC- LR	2.42 ± 0.18	5	0.91
QMI-GBR	3.28 ± 0.32	5	0.49
QMIC-GBR	2.82 ± 0.11	5	0.58

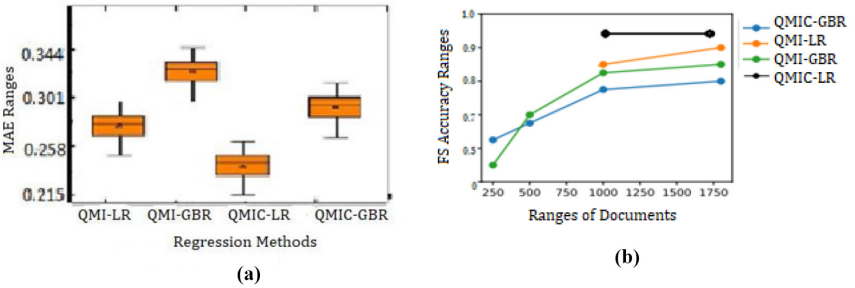

Figure 1

(a) Correlation Method MAE for all linear regression method with labeled x axis as linear regression methods and y axis as range. (b) FS accuracy for all linear regression method with constant $k = 5$.

stated baselines (Complete feature) over the Recursive feature elimination algorithm. Figure 1 is the sample output of the plot for accuracy and MAE for all linear regression methods keeping k constant with the value of 5.

5. Conclusion and Future Scope

A marking study that leverages quantum-aided function in the realm of machine learning training and testing process. An algorithm with the model of implementation is presented to determine the size of subsets k for the feature selection process with the meaning of two parameters σ and ϕ. However, QUBO formulation has a good impact on quantum annealing for feature selection there are some limitations in annealing like only a few problems are good candidates for quantum annealing. Classically all the datasets can be used for classifications, clustering, and ranking while in the realm of quantum annealing, effective results are not shown for all datasets. Meanwhile, for the ranking and clustering quantum computation supports a lot in giving acceptable results which can enrich the development of a ranking model that is inspired by feature selection and can achieve an effective quantum clustering or classification algorithm.

Meanwhile, as research is a continuous task with different compatibilities and sometimes demerits we have to proceed to the conclusion, which can be stated as the quantum-aided algorithm for feature selection works better for a specific dataset of text classification. In upcoming research, we can use the model for designing of efficient ranking algorithm in the realm of quantum computation.

References

[1] Chen, P.S., A faceted approach to information retrieval. *Journal of Information Optimization Sciences*, 29(4): p. 631-658 (2008).

[2] Gupta, A. and S.A. Begum, Efficient multi-cluster feature selection on text data. *Journal of Information Optimization Sciences*, 40(8): p. 1583-1598 (2019).

[3] Arora, I. and A. Saha, ELM and KELM based software defect prediction using feature selection techniques. *Journal of Information Optimization Sciences*, 40(5): p. 1025-1045 (2019).

[4] Tang, H.-C., T.-L. Chao, and K.-H. Hsieh, A weighted ranking function for ranking triangular fuzzy numbers. *Journal of Information Optimization Sciences*, 33(1): p. 149-158 (2012).

[5] Marchesin, S., A. Purpura, and G. Silvello, Focal elements of neural information retrieval models. An outlook through a reproducibility study. *Information Processing Management*, 57(6): p. 102109 (2020).

[6] Bender, E.M., T. Gebru, A. McMillan-Major, and S. Shmitchell. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? in Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (2021).

[7] Jain, A. and V. Jain, Sentiment classification of twitter data belonging to renewable energy using machine learning. *Journal of Information Optimization Sciences*, 40(2): p. 521-533 (2019).

[8] Kotsiantis, S.B., D. Kanellopoulos, and P.E. Pintelas, Data preprocessing for supervised leaning. *International Journal of Computer Science*, 1(2): p. 111-117 (2006).

[9] Pearson, K., LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, *Dublin philosophical magazine journal of science*, 2(11): p. 559-572 (1901).

[10] Ng, A.Y., On feature selection: learning with exponentially many irreverent features as training examples, Massachusetts Institute of Technology (1998).

[11] Blum, A.L. and P. Langley, Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1-2): p. 245-271 (1997).

[12] Motoda, H. and H. Liu, Feature selection, extraction and construction. *Communication of IICM*, 5(67-72): p. 2 (2002).

[13] Peng, H., F. Long, and C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-

redundancy. *IEEE Transactions on pattern analysis machine intelligence*, 27(8): p. 1226-1238 (2005).

[14] Miah, M., S.J. Miah, and S. Venkatraman, Blockchain: At a glance idea for information science researchers. *Journal of Information Optimization Sciences*, 42(7): p. 1589-1624 (2021).

[15] Chandrashekar, G. and F. Sahin, A survey on feature selection methods. *Computers Electrical Engineering*, 40(1): p. 16-28 (2014).

[16] Jović, A., K. Brkić, and N. Bogunović. A review of feature selection methods with applications. in 2015 38th international convention on information and communication technology, electronics and microelectronics (MIPRO). *Ieee* (2015).

[17] Nguyen, X.V., J. Chan, S. Romano, and J. Bailey. Effective global approaches for mutual information based feature selection. in Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (2014).

[18] Kshirsagar, D. and P. Agrawal, A study of feature selection methods for android malware detection. *Journal of Information Optimization Sciences*, 43(8): p. 2111-2120 (2022).

[19] Alsahaf, A., N. Petkov, V. Shenoy, and G. Azzopardi, A framework for feature selection through boosting. *Expert Systems with Applications*, 187: p. 115895 (2022).

[20] Glover, F., G. Kochenberger, and Y. Du, Quantum Bridge Analytics I: a tutorial on formulating and using QUBO models. 4OR, 17: p. 335-371 (2019).

[21] Lucas, A., Ising formulations of many NP problems. *Frontiers in Physics*, 2: p. 5 (2014).

[22] Bauckhage, C., N. Piatkowski, R. Sifa, D. Hecker, and S. Wrobel. A QUBO Formulation of the k-Medoids Problem. in LWDA (2019).

[23] Willsch, D., M. Willsch, H. De Raedt, and K. Michielsen, Support vector machines on the D-Wave quantum annealer. *Computer Physics Communications*, 248: p. 107006 (2020).

[24] Carugno, C., M. Ferrari Dacrema, and P. Cremonesi, Evaluating the job shop scheduling problem on a D-wave quantum annealer. *Scientific Reports*, 12(1): p. 6539 (2022).

[25] Reshef, D., Y. Reshef, M. Mitzenmacher, and P. Sabeti, Equitability analysis of the maximal information coefficient, with comparisons. arXiv preprint arXiv:, (2013).

[26] Nagy, M. and S.G. Akl, Quantum computing: Beyond the limits of conventional computation. *The International Journal of Parallel, Emergent Distributed Systems*, 22(2): p. 123-135 (2007).

[27] Matsumori, T., M. Taki, and T. Kadowaki, Application of QUBO solver using black-box optimization to structural design for resonance avoidance. *Scientific Reports*, 12(1): p. 12143 (2022).

[28] Monburinon, N., P. Chertchom, T. Kaewkiriya, S. Rungpheung, S. Buya, and P. Boonpou. Prediction of prices for used car by using regression models. in 2018 5th International Conference on Business and Industrial Research (ICBIR). *IEEE* (2018).

[29] Stamatopoulos, N., D.J. Egger, Y. Sun, C. Zoufal, R. Iten, N. Shen, and S. Woerner, Option pricing using quantum computers. *Quantum,* 4: p. 291 (2020).

[30] Rodriguez-Lujan, I., C. Elkan, C. Santa Cruz Fernández, and R. Huerta, Quadratic programming feature selection. *Journal of Machine Learning Research*, (2010).

[31] Anand, R., D. Aggarwal, and V. Kumar, A comparative analysis of optimization solvers. *Journal of Statistics Management Systems,* 20(4): p. 623-635 (2017).

[32] Milne, A., M. Rounds, and P. Goddard, Optimal feature selection in credit scoring and classification using a quantum annealer. White Paper 1Qbit, (2017).

[33] Sharma, K.K., Quantum Adiabatic Feature Selection. arXiv preprint arXiv:.08732, (2019).

[34] Yarkoni, S., E. Raponi, T. Bäck, and S. Schmitt, Quantum annealing for industry applications: Introduction and review. Reports on Progress in Physics, (2022).

[35] Abdulghani, F.A. and N.A. Abdullah, Hybrid deep learning model for Arabic text classification based on mutual information. *Journal of Information Optimization Sciences*, 43(8): p. 1901-1908 (2022).

[36] Poonia, R.C. and M. Kalra, Bridging approaches to reduce the gap between classical and quantum computing. *Journal of Information Optimization Sciences*, 37(2): p. 279-283 (2016).

[37] Glover, F., G. Kochenberger, R. Hennig, and Y. Du, Quantum bridge analytics I: a tutorial on formulating and using QUBO models. *Annals of Operations Research*, 314(1): p. 141-183 (2022).

[38] Lo, C.-Y. and W.-C. Lee, The application of linear programming technologies to the optimization of conflict problems. *Journal of Information Optimization Sciences*, 26(3): p. 693-713 (2005).

[39] Mohanty, B. and T. Sahoo, Mutual information based objective model for assessment of visual quality. *Journal of Information Optimization Sciences*, 43(5): p. 1151-1166 (2022).

[40] Ross, B.C., Mutual information between discrete and continuous data sets. PloS one, 9(2): p. e87357 (2014).

[41] Singh, H., N. Misra, V. Hnizdo, A. Fedorowicz, and E. Demchuk, Nearest neighbor estimates of entropy. *American Journal of Mathematical Management Sciences*, 23(3-4): p. 301-321 (2003).

[42] Luedtke, A. and L. Tran, The generalized mean information coefficient. arXiv preprint arXiv:, (2013).

[43] Stefanov, S.M., On the solution of quadratic programming problem with a feasible region defined as a Minkowski sum of a compact set and finitely generated convex closed cone. *Journal of Information Optimization Sciences*, 39(6): p. 1223-1230 (2018).

[44] Freedman, D.A., Statistical models: theory and practice (2009) : cambridge university press.

[45] Breiman, L., J.H. Friedman, R. Olshen, and C. Stone, Classification Regression trees. Wadsworth, p. 23-32 (1984).

[46] Guyon, I., J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines. *Machine learning*, 46: p. 389-422 (2002).

[47] Kramer, O. and O. Kramer, Scikit-learn. Machine learning for evolution strategies, p. 45-53 (2016).

[48] Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, and V. Dubourg, Scikit-learn: Machine learning in Python. the *Journal of Machine Learning Research*, 12: p. 2825-2830 (2011).

[49] McKinney, W. Data structures for statistical computing in python. in Proceedings of the 9th Python in Science Conference (2010). Austin, TX.