QUBO Based Quantum Annealing for Feature Selection, Instance Selection and Clustering

Sumaiyah Zahid[1,*,†], Muhammad Atif Tahir[2,†] and Muhammad Rabeet Sagri[3,†]

[1]National University of Computer & Emerging Sciences, NUCES-FAST
[2]Institute of Business Administration (IBA), Karachi
[3]GC Cybersecurity Inc.

Abstract

In this work, we present a unified QUBO based framework for three QuantumCLEF 2026 tasks: feature selection, instance selection, and document clustering. For feature selection, we combine relevance, stability, coverage, and redundancy signals within a QUBO formulation to identify compact and informative feature subsets. For instance selection, we propose a noise aware framework that integrates GMM based reliability estimation, predictive uncertainty, and local structural information to select representative training instances. For clustering, we introduce a query aware approach that combines relevance, density, and coverage signals with query biased farthest point sampling. Optimization is performed using both Simulated Annealing (SA) and Quantum Annealing (QA). For information retrieval, QA achieves an nDCG@10 score of 0.4472 using only 18 features. In recommender systems, SA achieves 0.0229 on the ICM-100 dataset using 75 features, while QA achieves 0.0307 on the ICM-400 dataset using 150 features. For instance selection, QA attains a Macro-F1 score of 58.5 while reducing the training data by 60.98%. For clustering, the proposed method achieves its best nDCG@10 score of 0.5901 for 10 centroids and a nDCG@10 score of 0.5263 for 25 centroids using SA, while QA achieves an nDCG@10 score of 0.5461 using 50 centroids. Across all tasks, QA provides substantially faster optimization while maintaining competitive performance, demonstrating the practical potential of QUBO based quantum optimization for information retrieval and machine learning applications.

Keywords

Feature Selection, Instance Selection, Clustering, Simulated Annealing, Quantum Annealing

1. Introduction

Recent developments in quantum computing have shown the growing interest in quantum inspired optimization methods to solve complex problems in machine learning, specifically feature selection, instance selection, and clustering. A prominent approach reformulates these combinatorial problems as Quadratic Unconstrained Binary Optimization (QUBO) models, which can be solved efficiently using Quantum Annealing (QA). This paradigm has attracted particular attention within the information retrieval and recommendation systems communities, where the QuantumCLEF initiative provides a systematic evaluation framework for quantum computing methods applied to retrieval and recommendation tasks [1, 2]. Despite promising results, the practical potential and current limitations of these approaches remain active areas of investigation.

QUBO has emerged as a popular modeling framework because machine learning objectives can be represented using binary decision variables. Earlier Mücke et al. [3] proposed a quantum feature selection framework, while Hellstern et al. [4] demonstrated its feasibility under current quantum hardware constraints. Further extensions include collaborative filtering based feature selection for recommendation systems by Anelli et al. [5] and mutual information driven formulations by Pranjic et al. [6]. Recent studies have also explored quantum SVM based feature selection and feature mapping strategies [7], as well as general QUBO formulations for model training and feature optimization [8].

CLEF 2026 Working Notes, September 21-24, 2026, Jena, Germany
*Corresponding author.
†These authors contributed equally.
sumaiyah@nu.edu.pk (S. Zahid); atiftahir@iba.edu.pk (M. A. Tahir); rabeet@ghangorcloud.com (M. R. Sagri)
0000-0002-7138-8482 (S. Zahid); 0000-0003-1366-8408 (M. A. Tahir); 0000-0002-9421-8566 (M. R. Sagri)


Earlier annealing based studies for a feature selection approach includes Simulated Annealing approaches [9] and foundational QA formulations [10, 11], which establish QA as a general combinatorial optimization framework.

Instance selection follows a similar paradigm at the data level, where the objective is to reduce training cost while preserving performance. Pasin et al. [12] demonstrated that quantum annealing can be effectively used for transformer fine tuning through QUBO based subset selection, while Cunha et al. [13] proposed scalable confidence based instance selection for large scale NLP systems. Related work by Nembrini et al. [14] and Borle et al. [15] further supports the effectiveness of quantum and hybrid subset selection methods in both feature and instance domains.

Major studies shown that classical clustering objectives can be reformulated as QUBO problems by QA. Kurihara et al. [16] establishes the idea of clustering via quantum annealing, followed by Bauckhage et al. [17], who formalized k-medoids as a QUBO optimization problem. Matsumoto et al. [18] extended this to distance based clustering formulations, while Zaech et al. [19] proposed balanced clustering using adiabatic quantum computing. More recently, Alvarez-Girón et al. [20] demonstrated QA based clustering in real world CLEF evaluation settings. Practical implementation is supported by the D-Wave Ocean framework [21], while broader applications such as optimal power flow optimization [22] highlight the general applicability of QA to largescale combinatorial problems.

The above studies show that feature selection, instance selection, and clustering can be naturally unified under the QUBO framework and efficiently addressed using QA. Across these domains, QA consistently demonstrates compact representations and competitive performance while reducing computational cost. However, challenges such as scalability, embedding overhead, and hardware noise remain open problems, which opens further research areas into hybrid quantum classical approaches for large scale machine learning applications. The further sections of the paper are organized as follows. Section 2 presents the proposed methodology and describes the QUBO based formulations developed for the three QuantumCLEF tasks including the feature selection framework for information retrieval and recommender systems, the noiseaware instance selection framework, and the query aware clustering methodology. For each task, we discuss the feature engineering process, scoring mechanisms, QUBO formulation, optimization strategy using Simulated Annealing (SA) and Quantum Annealing (QA), and evaluation metrics. Section 3 presents and analyzes the experimental results obtained across all tasks, highlighting the trade offs between effectiveness, dimensionality reduction, and optimization cost. Finally, Section 4 concludes the paper, summarizes the key findings, and outlines potential directions for future research in quantum inspired optimization for machine learning and information retrieval applications.

2. Methodology

2.1. Task 1A – Feature Selection: Information Retrieval

Given a MQ2007 learning-to-rank dataset with document feature vectors, relevance labels, and query identifier. The objective is to select a subset of features, that maximizes ranking performance when used to train a learning-to-rank model evaluated using nDCG@10.

2.1.1. Mutual Information with Stability Estimation

Feature relevance is estimated using Mutual Information (MI), which captures both linear and non linear dependencies between a feature and the target variable. A KNN-based estimator is employed to handle continuous feature spaces, and MI is computed across multiple GroupKFold splits to preserve query level grouping. Mutual Information score of feature i computed on fold f is represented as $MI_i^{(f)}$. For each feature, MI scores are averaged across folds to obtain a robust relevance estimate, while the standard deviation is used to assess stability. Features exhibiting high variability across folds are penalized, resulting in a stability-aware importance score s_i that combines both relevance and consistency.

2.1.2. Redundancy Modeling

Feature redundancy is modeled using the absolute Pearson correlation between any two features i and j, denoted as r_{ij}. To encourage diversity in the optimization process, only feature pairs with correlation values greater than a predefined threshold are considered redundant and penalized in the objective function. In this study, the threshold was set to 0.15. This value was selected after experimenting with several correlation thresholds and evaluating their impact on feature selection performance on validation split.

2.1.3. QUBO Formulation

We formulate the feature selection problem as a QUBO over binary decision variables $x_i \in \{0, 1\}$, where $x_i = 1$ indicates that feature i is selected. The objective function is composed of three components and defined as:

$$\min_{x \in \{0,1\}^d} \left[-\sum_{i=1}^{d} s_i x_i + \delta \sum_{i<j} r_{ij} x_i x_j + \lambda \left(\sum_{i=1}^{d} x_i - k \right)^2 \right]$$

The first term promotes the selection of informative and relevant features based on their relevance scores. The second term introduces a redundancy penalty that discourages the simultaneous selection of highly correlated features, thereby maintaining diversity within the selected subset. The third term is a cardinality constraint that enforces the selection of exactly k features by penalizing deviations from the desired subset size.

2.1.4. Evaluation Metric

To evaluate the quality of the selected feature subsets, a probabilistic ranking model is trained using only the selected features. Logistic regression is used to estimate relevance probabilities in the reduced feature space. The model assigns higher scores to more relevant items based on the learned linear decision function. For each query, the model produces a ranked list of items based on predicted relevance scores, and the final performance is obtained by averaging nDCG@10 across all queries.

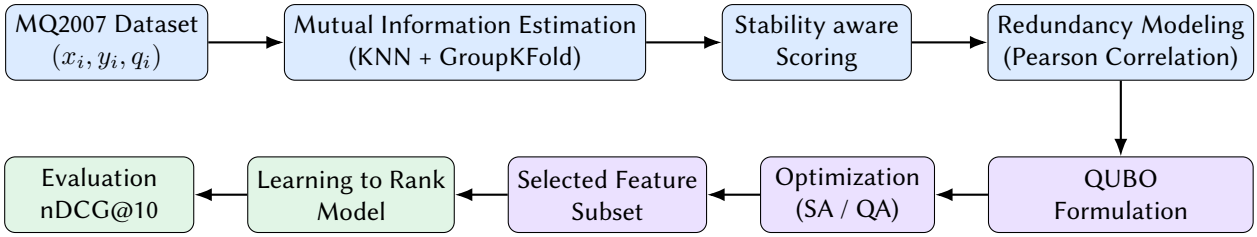

Figure 1: Flowchart of Task 1A - Feature selection : Information Retreival

2.2. Task 1B – Feature Selection: Recommender Systems

We submitted four runs in total, comprising two configurations for the 100-ICM setting and two for the 400-ICM setting. For each setting, we evaluate both a simulated annealing based solver and a quantum annealing based solver to study the effect of different optimization paradigms on feature selection performance.

We propose a sparse feature selection framework for recommendation systems based on Quadratic Unconstrained Binary Optimization (QUBO). The method aims to identify a compact and informative subset of item-content features that maximizes recommendation effectiveness while minimizing redundancy among features. Furthermore, the formulation is designed to be compatible with quantum annealing hardware constraints, enabling efficient embedding and optimization on QPU based architectures.

2.2.1. Feature Score

Let $\mathbf{R}$ denote the user item interaction matrix and $\mathbf{F}$ denote the Item Content Matrix (ICM). To reduce optimization complexity, an initial feature preselection stage is performed. In this stage, item popularity is computed based on user interactions, and feature importance is estimated using the item content matrix denoted by s_i. The top-150 highest scoring features are retained to form the reduced feature matrix $\mathbf{F}_r$, out of which, four feature quality signals are computed: collaborative relevance, separability, coverage, and redundancy.

1. **Collaborative relevance:** represented as C_i, measures the association between a feature and frequently interacted items.
2. **Feature separability:** represented as S_i, captures the ability of a feature to differentiate between popular items and less popular items.
3. **Coverage:** represented as V_i, captures how well a feature represents structurally important items by aggregating item level contributions.
4. **Feature redundancy:** represented as R_i, is estimated using the co-occurrence matrix $\mathbf{A} = \mathbf{F}_r^T \mathbf{F}_r$. The redundancy score for each feature is then computed by summing its associations with all other features..

All signals are normalized prior to optimization. Final composite feature score was defined as q_i. The parameters α, β, γ, δ, and λ control the contributions of collaborative relevance, separability, coverage, redundancy suppression, and the cardinality constraint, respectively. To determine suitable parameter values, a grid search strategy was employed over multiple candidate combinations. Each parameter configuration was evaluated on a local validation split, and the combination yielding the best validation performance was selected for the final submission. The selected parameters were then used to compute the composite feature score q_i.

$$q_i = \alpha C_i + \beta S_i + \delta V_i - \gamma R_i,$$

2.2.2. QUBO Formulation

We formulate the feature selection problem as a QUBO over binary decision variables $x_i \in \{0, 1\}$, where $x_i = 1$ indicates that feature i is selected. The objective function is composed of three components and defined as:

$$\min_{\mathbf{x}} \left(-\sum_i q_i x_i + \sum_{i<j} w_{ij} x_i x_j + \lambda \left(\sum_i x_i - k \right)^2 \right),$$

where w_{ij} represents sparse redundancy penalties derived from feature co-occurrence.
The first term promotes the selection of informative and relevant features based on their relevance scores. The second term introduces a redundancy penalty that discourages the simultaneous selection of highly correlated features, thereby maintaining diversity within the selected subset. The third term is a cardinality constraint that enforces the selection of exactly k features by penalizing deviations from the desired subset size.

2.2.3. Evaluation Metric

The selected features are used to construct a reduced Item Content Matrix for the Item based KNN Content Based Filtering recommender. The recommender is trained using cosine similarity with neighborhood truncation and shrinkage regularization. Performance is evaluated on a validation split using nDCG@10. This evaluation measures ranking quality while validating the effectiveness of the quantum optimized feature subset for recommendation tasks.

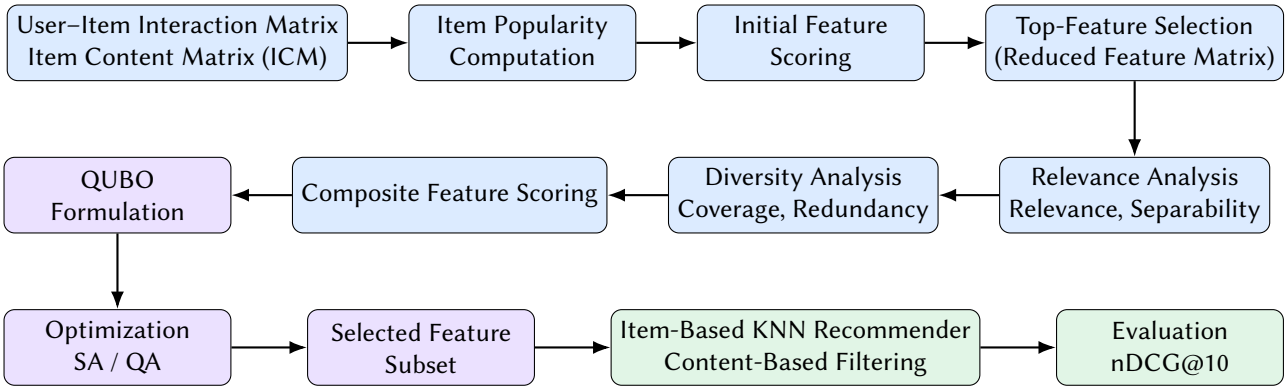

Figure 2: Flowchart of Task 1B - Feature selection : Recommender System

2.3. Task 2 – Instance Selection

We propose a quantum annealing based framework for selecting a representative and noise robust subset of training instances for text classification, reducing the cost of fine tuning large language models (e.g., Llama3.1) while maintaining comparable performance to full data training. Given that the dataset contains around 20% label noise, we formulate the problem as a Quadratic Unconstrained Binary Optimization (QUBO) model and solve it using Simulated Annealing (SA) and Quantum Annealing (QA) to obtain a compact subset that balances representativeness and label quality.

2.3.1. Learning Signal and GMM Based Noise Estimation

Given a set of embedding vectors $X = x_1, x_2, \ldots, x_n$ with labels y, we estimate several signals that characterize the reliability and representativeness of each training instance. A logistic regression classifier is first trained on the embedding space to obtain class probabilities. The negative log likelihood loss was computed to fit a Gaussian Mixture Model (GMM) in order to separate clean and noisy samples. The reliability of each training instance is estimated by combining the probability of belonging to the low loss component identified by the GMM with the consistency of its label within the local neighborhood.

$$R(x_i) = P_{\mathrm{GMM}}(x_i), L(x_i),$$

where $P_{\mathrm{GMM}}(x_i)$ denotes the probability that instance x_i belongs to the clean component and $L(x_i)$ measures label agreement within its local neighborhood.'
Predictive uncertainty is estimated using the normalized entropy of the classifier outputs, where the normalization is performed with respect to the total number of classes C. Higher uncertainty values indicate lower confidence in the predicted class probabilities.

$$U(x_i) = -\frac{1}{\log C} \sum_{c=1}^{C} p_{ic} \log p_{ic},$$

A cosine k-nearest-neighbor graph is constructed to capture local structure in the embedding space. Density and redundancy scores are then derived from the distances between each instance and its neighboring samples. In this context, $\mathcal{N}_i$ represents the set of K nearest neighbors of x_i.

$$E(x_i) = \frac{1}{K} \sum_{j \in \mathcal{N}_i} d_{\cos}(x_i, x_j), \quad D(x_i) = \frac{1}{E(x_i)}.$$

2.3.2. Divide-and-Mix Composite Scoring

The estimated signals are grouped into reliability, uncertainty, and structural components and combined through a divide-and-mix strategy. The final sample quality score is defined as

$$S(x_i) = \alpha R(x_i) - \beta U(x_i) + \gamma D(x_i) - \delta E(x_i),$$

where α, β, γ, and δ control the weightages of reliability, uncertainty, density, and redundancy, respectively. Higher scores indicate informative, representative, and noise resistant training instances.

2.3.3. QUBO Formulation

To make the optimization process feasible on current quantum hardware, the dataset is first divided into class balanced batches. A subset selection budget is then assigned to each batch in proportion to its size, ensuring that all classes remain fairly represented while still achieving the desired overall data reduction. Within each batch, binary decision variables $x_i \in \{0, 1\}$ are used to indicate whether a sample d_i is selected. The subset selection task is then formulated as the following QUBO optimization problem:

$$\min_{\mathbf{x}} \left(-\alpha \sum_i S(d_i) x_i + \beta \sum_{i<j} \mathrm{sim}(d_i, d_j) x_i x_j + \lambda \left(\sum_i x_i - k_b \right)^2 \right),$$

where $S(d_i)$ denotes the composite quality score of sample d_i, $\mathrm{sim}(d_i, d_j)$ represents the cosine similarity between samples d_i and d_j, and k_b is the number of samples to be selected from batch b. The parameters α, β, and λ control the influence of sample quality, diversity regularization, and the cardinality constraint, respectively.

The first term favors high quality samples, the second penalizes selecting highly similar samples to encourage diversity, and the third enforces the required subset size. The resulting Binary Quadratic Model (BQM) is solved using either SA or QA. Each batch specific QUBO is optimized independently, and the selected samples are aggregated across all batches. Duplicate selections are removed, and if the final subset exceeds the budget k, a trimming step is applied to retain the highest-scoring instances.

2.3.4. Evaluation Metrics

The selected subset is evaluated in a downstream setting by fine-tuning a learning model and reporting the Macro F1 score on a held out test set. Meanwhile, reduction is evaluated based on the degree of dataset compression achieved.

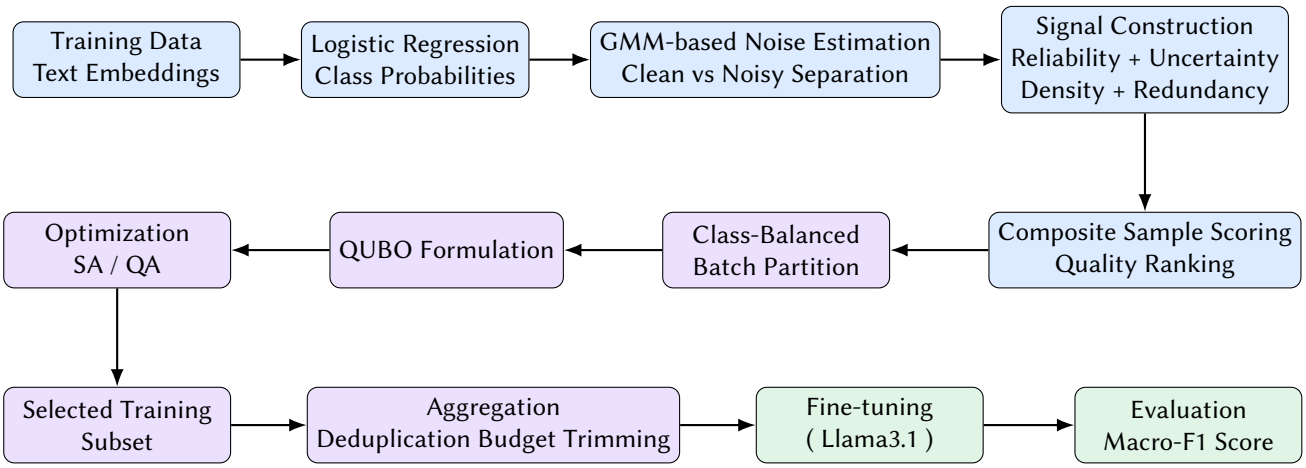

Figure 3: Flowchart of Task 2: Instance selection

2.4. Task 3 – Clustering

In this task, we address the problem of document clustering using both Simulated Annealing (SA) and Quantum Annealing (QA). The proposed framework formulates clustering as a constrained Quadratic Unconstrained Binary Optimization (QUBO) problem and integrates query aware signals to improve retrieval oriented clustering quality. The proposed approach consists of two major stages:

1. Query aware document scoring
2. Query biased Farthest Point Sampling (FPS) for pivot reduction

2.4.1. Query aware document scoring

Document $\mathcal{D}$ and query embeddings are loaded and L2-normalized to enable cosine similarity computations. A relevance weighted query signal is used to incorporate retreieval relevance into clustering. The relevance score $R(d_i)$ for each document d_i is calculated using query embeddings q_j and their relative weights w_j.

$$R(d_i) = \sum_{j=1}^{m} w_j \cdot \cos(d_i, q_j)$$

To identify representative centroids, local density is computed using the average cosine similarity to the nearest neighbors of each document. Dense regions indicate semantically important areas in the embedding space. For each document, the density score is defined as the mean similarity to its k nearest neighbors, where nn_j denotes the j-th nearest neighbor of document d_i.

$$D(d_i) = \frac{1}{k} \sum_{j=1}^{k} \cos(d_i, nn_j)$$

A composite quality score is first constructed to evaluate document importance by combining query relevance and local density:

$$S(d_i) = \alpha R(d_i) + \beta D(d_i), \quad d_i \in \mathcal{D}$$

where α and β are weighting hyperparameters controlling the influence of retrieval relevance and structural density respectively. The scores are normalized to the range $[0, 1]$ before further processing.

2.4.2. Query Biased Farthest Point Sampling

Since solving the clustering problem on the full dataset is computationally expensive for current quantum hardware, a pivot reduction strategy based on Farthest Point Sampling (FPS) is employed. Unlike traditional FPS, the proposed method incorporates the composite score $S(d_i)$ to prioritize query aware and representative documents. At each iteration, the next pivot is selected as:

$$p^* = \arg\max_{i \in \mathcal{D}} \left(d_i^{\min} (1 + S(d_i)) \right)$$

where $d_i^{\min}$ denotes the minimum cosine distance between document d_i and the currently selected pivots.
After selecting the pivot set $\mathcal{P} \subset \mathcal{D}$, a coverage score is computed to capture how well each pivot represents the dataset. It is defined as:

$$C(p_j) = \sum_{i=1}^{N} \mathbb{I}\left(\arg\min_{p_k \in \mathcal{P}} d_{\cos}(x_i, x_{p_k}) = p_j \right)$$

where $\mathcal{P}$ denotes the selected pivot set and $d_{\cos}(\cdot, \cdot)$ represents cosine distance.
The final pivot level scoring function is defined only over the selected pivots as:

$$S(p_j) = \alpha R(p_j) + \beta D(p_j) + \gamma C(p_j), \quad p_j \in \mathcal{P}$$

where $R(p_j)$, $D(p_j)$, and $C(p_j)$ denote the relevance, density, and coverage scores for pivots, respectively, and α, β, and γ control their relative contributions.

2.4.3. QUBO Formulation

The pivot selection problem is formulated as a QUBO over a candidate set of pivots $\mathcal{P}$ obtained via query biased Farthest Point Sampling. Let $x_i \in \{0, 1\}$ indicate whether pivot p_i is selected, with each pivot assigned a score s_i. The objective selects exactly k pivots while balancing quality and diversity:

$$\min_{x \in \{0,1\}^{|\mathcal{P}|}} \left[-\sum_{p_i \in \mathcal{P}} s_i x_i + \lambda \sum_{i<j} \text{sim}(p_i, p_j)\, x_i x_j + A \left(\sum_{p_i \in \mathcal{P}} x_i - k \right)^2 \right].$$

The first term promotes high quality pivots, the second term penalizes redundancy via pairwise similarity, and the third term enforces selection of exactly k pivots. The resulting QUBO is solved using simulated or quantum annealing, and the selected pivots are then used as centroids to cluster documents via nearest centroid assignment in embedding space.

2.4.4. Evaluation Metrics

The quality of the resulting clusters is evaluated using two complementary metrics:

- **Davies-Bouldin Index (DBI):** A clustering quality measure that evaluates intra cluster compactness and inter cluster separation. Lower values indicate more well formed and distinct clusters.
- **nDCG@10:** A ranking ased retrieval metric that assesses how effectively relevant documents are retrieved within the top-10 results of the nearest cluster for each query. Higher values indicate better retrieval performance.

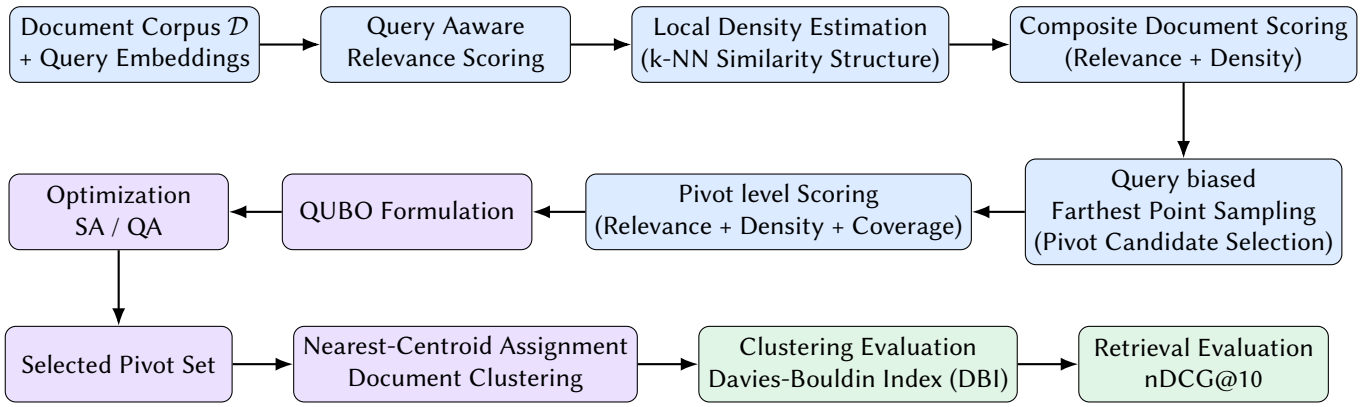

Figure 4: Flowchart of Task 3: Clustering.

3. Results

The results in Table 1 show that both QA and SA achieve competitive ranking performance while using substantially fewer features than the baseline. Although the baseline attains the highest nDCG@10 score of 0.4473 using 46 features, QA achieves a nearly identical score of 0.4472 with only 18 features. Similarly, SA obtains an nDCG@10 of 0.4412 using just 12 features. The larger subset selected by QA suggests that the quantum annealing process favored retaining a small number of additional informative features, leading to a marginal improvement in ranking effectiveness compared with the more compact subset identified by SA. These results demonstrate that QA and SA can maintain ranking effectiveness while significantly reducing feature dimensionality, with QA providing the best balance between performance and feature reduction. Figure 5 shows that QA and SA share 6 common features while selecting 13 and 7 unique features, respectively. Figure 6 illustrates the feature wise agreement across all feature indices, explicitly indicating for each feature whether it was selected by QA only, SA only, both algorithms, or neither.

Table 1
Result of Task 1A

Submission ID	nDCG@10	Annealing Time (us)	Type	Features
1A_MQ2007_SA_FAST-NUCES_k12	0.4412	4322921	S	12
1A_MQ2007_QA_FAST-NUCES_k12	0.4472	154243	Q	18
BASELINE_ALL	0.4473	-	-	46
RFE_BASELINE_HALF	0.4450	-	-	23

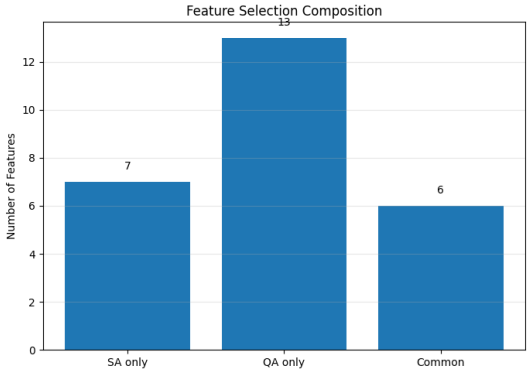

Figure 5: Feature selection comparison for MQ2007 dataset

Figure 6: Feature Wise Agreement of MQ2007 dataset

From the results shown in Tables 2 and 3, the effectiveness of the proposed framework based on QA and SA techniques can be seen in different feature dimension settings (ICM 100 and ICM 400). In both settings, the best performance of the baseline technique occurs when using the complete set of features with scores of nDCG@10 of 0.0226 for 100 features and 0.0328 for 400 features. In the case of ICM 100 setting, the SA approach provides an enhanced nDCG@10 score of 0.0229 as compared to the baseline technique of 0.0226 by reducing the feature set from 100 to 75. The QA technique also gives a similar performance level of 0.0218 when using 72 features. However, the use of QA reduces annealing time as compared to SA.

Figure 7 shows that QA and SA share 55 common features, while selecting 18 and 21 unique features, respectively. Figure 8 presents the feature wise agreement across all feature indices, indicating both consensus on important features and complementary feature selection patterns between the two methods.

Table 2
Result of Task 1B - ICM 100 dataset

Submission ID	nDCG@10	Annealing Time (us)	Type	Features
1B_100_ICM_SA_FAST-NUCES_Ensemble	0.0229	15485673	S	75
1B_100_ICM_QA_FAST-NUCES_Ensemble	0.0218	456086	Q	72
ALL_FEATURES	0.0226	-	-	100

Table 3
Result of Task 1B - ICM 400 dataset

Submission ID	nDCG@10	Annealing Time (us)	Type	Features
1B_400_ICM_SA_FAST-NUCES_Ensemble	0.0289	129408361	S	200
1B_400_ICM_QA_FAST-NUCES_Ensemble	0.0307	36735	Q	150
ALL_FEATURES	0.0328	-	-	400

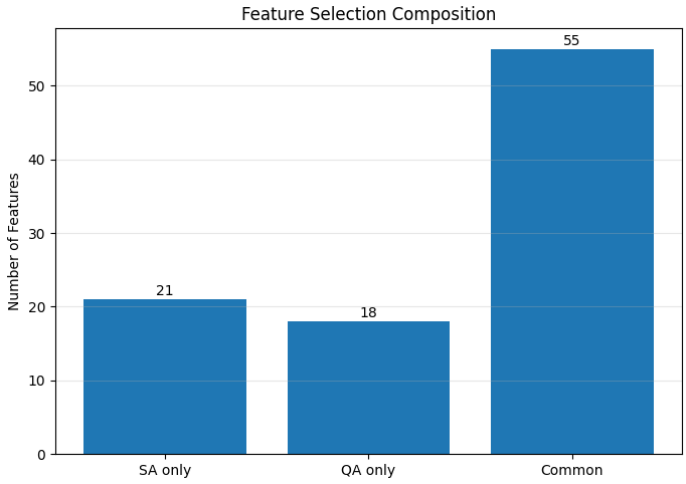

Figure 7: Feature Selection comparison of ICM 100 dataset.

For larger ICM 400 setting, the performance of the QA technique is better than that of the SA technique with the nDCG@10 scores being 0.0307 and 0.0289, respectively using 150 and 200 features. Despite of both the approaches, neither can beat the baseline performance in terms of using the complete feature set. Figure 9 shows that QA and SA share 92 common features, while selecting 59 and 109 unique features, respectively. Figure 10 presents the feature wise agreement across all feature indices, indicating both consensus on important features and complementary feature selection patterns between the two methods.

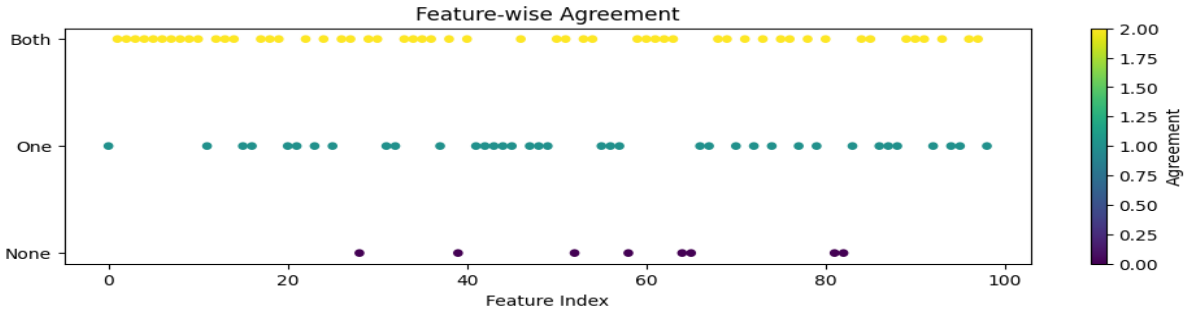

Figure 8: Feature Wise Agreement of ICM 100 dataset.

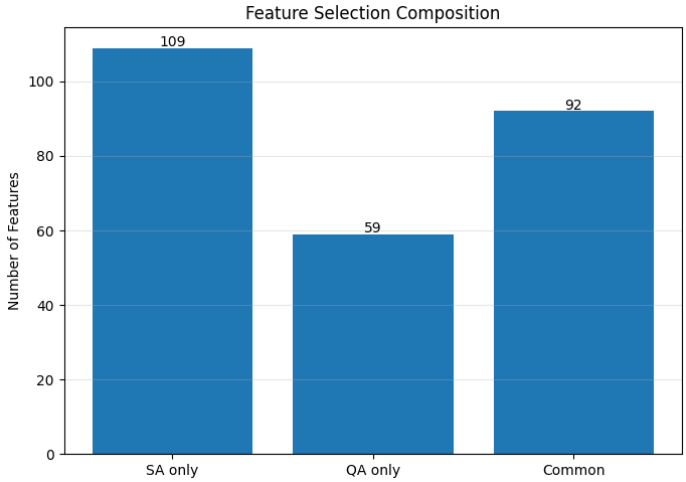

Figure 9: Feature Selection comparison of ICM 400 dataset.

Figure 10: Feature Wise Agreement of ICM 400 dataset.

The results in Table 4 evaluate performance on Task 2 under noisy and clean settings in terms of Macro F1 score, feature reduction, and computational cost. While the baselines achieve higher Macro F1 scores (88.9 in noise free and 84.3 in noisy settings), they rely on the full feature space and incur higher computational cost. In contrast, our methods significantly improve feature efficiency: Algorithm 1 achieves a Macro F1 of 57.9 with 75.03% feature reduction and lower runtime (541.7 s), while DivideMix (Simulated Annealing) obtains a Macro F1 of 56.3 with 70.04% reduction, at the expense of higher annealing time. Notably, Quantum Annealing based DivideMix achieves the best overall trade off, obtaining a Macro F1 score of 58.5 with 60.98% feature reduction and lower annealing time than the

SA based variant, which is significantly more computationally expensive. Overall, while baselines achieve higher absolute accuracy, our framework offers a better efficiency–performance balance through substantial feature reduction and reduced inference cost, with Quantum Annealing further improving scalability over Simulated Annealing.

Table 4
Result of Task 2

Submission ID	Macro F1 Avg	Avg Reduction	Avg Fine-Tuning Prediction Time (s)	Avg Annealing Time (us)	Type
FAST-NUCES_Algorithm1	57.9(8.8)	75.03%	541.7(0.3)	167180028	S
FAST-NUCES_DivideMix	56.3(9.1)	70.04%	642.4(28.6)	2918988448	S
FAST-NUCES_DivideMix	58.5(11.9)	60.98%	797.4(11.1)	2643107	Q
Baseline_without_noise	88.9(0.8)	-	1997.3(5.7)	-	-
Baseline_with_noise	84.3(2.5)	-	1904.3(1.4)	-	-

The results in Tables 5, 6, and 7 are reported for 10, 25, and 50 centroids in terms of nDCG@10, DBI, and computational cost. In general, our approach consistently improves clustering quality over the baseline. Across all settings, query-aware methods, particularly FPS QUERYAWARE, achieve the strongest overall performance. For 10 centroids, KMEDOIDS and KMEDOIDS (QA) achieve the best nDCG@10 of 0.5901 compared to the baseline score of 0.5509, while FPS QUERYAWARE improves DBI to 6.6164 from 7.9892. For 25 centroids, the baseline achieves an nDCG@10 of 0.5284, while KMEDOIDS (0.5263) and FPS QUERYAWARE (0.5173) remain competitive with improved clustering structure. For 50 centroids, FPS QUERYAWARE achieves the best results, reaching 0.5461 for QA and 0.5411 for SA in nDCG@10 compared to the baseline score of 0.4656, along with improved DBI values of 4.3671 (QA) and 4.5136 (SA). These results suggest that incorporating query relevance into pivot selection helps identify more representative centroids, leading to improved retrieval effectiveness and cluster quality. Additionally, the gains become more pronounced as the number of centroids increases, highlighting the scalability of the proposed approach. In all settings, Quantum Annealing significantly reduces optimization time compared to Simulated Annealing while maintaining comparable or better clustering performance.

Table 5
Results of Task 3 for 10 Centroids

Submission ID	Centroids	nDCG@10	DBI	Annealing Time (us)	Type
FAST-NUCES_KMEDOIDS	10	0.5489	6.9336	2277560	S
FAST-NUCES_KMEDOIDS_1	10	0.5901	7.2754	7881661	S
FAST-NUCES_KMEDOIDS_QA	10	0.5901	7.2754	7979288	S
FAST-NUCES_FPS-QUERYAWARE	10	0.5709	6.6164	11819054	S
FAST-NUCES_FPS-QUERYAWARE	10	0.5497	7.0623	16035	Q
Baseline_10	10	0.5509	7.9892	-	-

Table 6
Results of Task 3 for 25 Centroids

Submission ID	Centroids	nDCG@10	DBI	Annealing Time (us)	Type
FAST-NUCES_KMEDOIDS	25	0.4592	6.9607	182119357	S
FAST-NUCES_KMEDOIDS_1	25	0.5263	5.7743	7766515	S
FAST-NUCES_KMEDOIDS_QA	25	0.4804	5.8211	7798667	S
FAST-NUCES_FPS-QUERYAWARE	25	0.5173	5.4729	14502851	S
FAST-NUCES_FPS-QUERYAWARE	25	0.4793	5.6272	16028	Q
Baseline_25	25	0.5284	6.1201	-	-

Table 7
Results of Task 3 for 50 Centroids

Submission ID	Centroids	nDCG@10	DBI	Annealing Time (us)	Type
FAST-NUCES_KMEDOIDS	50	0.5052	6.1287	182210944	S
FAST-NUCES_KMEDOIDS_1	50	0.4661	5.8774	7978485	S
FAST-NUCES_KMEDOIDS_QA	50	0.5158	6.0449	7812281	S
FAST-NUCES_FPS-QUERYAWARE	50	0.5411	4.3671	18007884	S
FAST-NUCES_FPS-QUERYAWARE	50	0.5461	4.5136	16045	Q
Baseline_50	50	0.4656	5.3679	-	-

4. Conclusion

Our work demonstrates that we provide a simple and effective approach for feature selection, instance selection, and clustering within a single QUBO based framework. The main idea is to select a less but useful set of features by balancing relevance, redundancy, and selection size, while also combining features in a way that preserves important information. In Task 1, our approach achieves strong compression without losing the overall structure of the data. In Task 2, it remains stable under both noisy and clean conditions, giving competitive Macro F1 scores while greatly reducing the number of features and overall cost. Task 3 further supports our approach, where query aware selection improves clustering quality across different centroid settings, and QA consistently reduces optimization time compared to SA. Overall, the results across all tasks show a clear and practical trade off between efficiency and performance, highlighting that QUBO based optimization with quantum-inspired methods can scale well while still maintaining good quality results.

Code Availability

The source code, experimental pipelines, and implementation details used in this study are publicly available at: https://github.com/SumaiyahZahid/QCLEF-2026.

Acknowledgements

We thank the Quantum CLEF 2026 organizers and the D-Wave team for providing access to quantum annealing hardware.

Declaration on Generative AI

During the preparation of this work, the author(s) used GPT-4 and Grammarly in order to: Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] A. Pasin, M. F. Dacrema, W. Cunha, M. A. Gonçalves, P. Cremonesi, N. Ferro, Overview of quantumclef 2026: The third quantum computing challenge for information retrieval and recommender systems at CLEF, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction - 17th International Conference of the CLEF Association, CLEF 2026, Jena, Germany, September 21-24, 2026, Proceedings, Lecture Notes in Computer Science, Springer, 2026.

[2] A. Pasin, M. F. Dacrema, W. Cunha, M. A. Gonçalves, P. Cremonesi, N. Ferro, Quantumclef 2026: Overview of the third quantum computing challenge for information retrieval and recommender

systems at CLEF, in: Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2026, Jena, Germany, 21-24 September 2026, CEUR Workshop Proceedings, CEUR-WS.org, 2026.
[3] S. Mücke, R. Heese, S. Müller, M. Wolter, N. Piatkowski, Feature selection on quantum computers, Quantum Machine Intelligence 5 (2023).
[4] G. Hellstern, V. Dehn, M. Zaefferer, Quantum computer based feature selection in machine learning, arXiv preprint arXiv:2306.10591 (2023).
[5] V. W. Anelli, A. Bellogín, T. Di Noia, A. Ferrara, Feature selection for recommender systems with quantum computing, Entropy 23 (2021) 1353.
[6] D. Pranjic, B. C. Mummaneni, K. Tutschku, Quantum annealing based feature selection in machine learning, arXiv preprint arXiv:2411.19609 (2024).
[7] S. Zahid, M. A. Tahir, Unlocking quantum svm potential: optimal feature map generation and feature selection, Physica Scripta 100 (2025) 015120. URL: https://doi.org/10.1088/1402-4896/ad9e39. doi:10.1088/1402-4896/ad9e39.
[8] P. Date, D. Arthur, L. Pusey-Nazzaro, Qubo formulations for training machine learning models, Quantum Computing Applications 1 (2023) 100–112.
[9] M. S. Haque, M. Fahim, M. Ibrahim, An exploratory study on simulated annealing for feature selection in learning-to-rank, arXiv preprint arXiv:2310.13269 (2023).
[10] S. Mücke, R. Heese, S. Müller, M. Wolter, N. Piatkowski, Feature selection on quantum computers, Quantum Machine Intelligence 4 (2022) 1–11.
[11] S. Yarkoni, E. Raponi, T. Bäck, S. Schmitt, Quantum annealing for industry applications: Introduction and review, Reports on Progress in Physics 85 (2022) 104001.
[12] A. Pasin, W. Cunha, M. A. Gonçalves, N. Ferro, A quantum annealing instance selection approach for efficient and effective transformer fine-tuning, in: Proceedings of the 2024 ACM SIGIR International Conference on Theory of Information Retrieval, 2024, pp. 205–214.
[13] W. Cunha, C. França, G. Fonseca, L. Rocha, M. A. Gonçalves, An effective, efficient, and scalable confidence-based instance selection framework for transformer-based text classification, in: Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023, pp. 665–674.
[14] R. Nembrini, M. F. Dacrema, P. Cremonesi, Feature selection for recommender systems with quantum computing, Journal of Computing Frontiers 10 (2024) 45–57.
[15] N. Borle, N. Zecevic, et al., Feature selection with quantum annealing for interpretable and robust machine learning, Quantum Machine Intelligence 5 (2023) 1–15.
[16] K. Kurihara, S. Tanaka, S. Miyashita, Quantum annealing for clustering, in: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI), AUAI Press, 2009, pp. 317–324.
[17] C. Bauckhage, N. Piatkowski, R. Sifa, D. Hecker, S. Wrobel, A qubo formulation of the k-medoids problem, in: LWDA, 2019, pp. 54–63.
[18] K. Matsumoto, Y. Hamakawa, K. Tatsumura, K. Kudo, Distance-based clustering using qubo formulations, Scientific Reports 12 (2022) 2669. doi:10.1038/s41598-022-06559-z.
[19] J.-N. Zaech, M. Danelljan, T. Birdal, L. Van Gool, Probabilistic sampling of balanced k-means using adiabatic quantum computing, arXiv preprint arXiv:2310.12153 (2023).
[20] W. Alvarez-Girón, J. Téllez-Torres, J. Tovar-Cortés, H. Gómez-Adorno, Team qiimas on task 2 - clustering: Quantum annealing for k-medoids optimization, in: Working Notes of CLEF 2024 Conference and Labs of the Evaluation Forum, 2024. URL: https://bitbucket.org/eval-labs/qc24-qiimas/src/main/.
[21] D-Wave Systems Inc., Ocean software documentation, 2023. URL: https://docs.ocean.dwavesys.com/.
[22] T. Morstyn, Annealing-based quantum computing for combinatorial optimal power flow, IEEE Transactions on Smart Grid PP (2022) 1–1. doi:10.1109/TSG.2022.3200590.