



 Latest updates: <https://dl.acm.org/doi/10.1145/3705328.3748029>

SHORT-PAPER

Estimating Quantum Execution Requirements for Feature Selection in Recommender Systems Using Extreme Value Theory

JIAYANG NIU, RMIT University, Melbourne, VIC, Australia

QIHAN ZOU, University of Melbourne, Melbourne, VIC, Australia

JIE LI, RMIT University, Melbourne, VIC, Australia

KE DENG, RMIT University, Melbourne, VIC, Australia

MARK SANDERSON, RMIT University, Melbourne, VIC, Australia

YONGLI REN, RMIT University, Melbourne, VIC, Australia

Open Access Support provided by:

University of Melbourne

RMIT University



PDF Download
3705328.3748029.pdf
25 January 2026
Total Citations: 0
Total Downloads: 2003

Published: 22 September 2025

[Citation in BibTeX format](#)

RecSys '25: Nineteenth ACM Conference
on Recommender Systems
September 22 - 26, 2025
Prague, Czech Republic

Conference Sponsors:
SIGCHI

Estimating Quantum Execution Requirements for Feature Selection in Recommender Systems Using Extreme Value Theory

Jiayang Niu

School of Computing Technologies
RMIT University
Melbourne, Australia
s4068570@student.rmit.edu.au

Qihan Zou

School of Mathematics and Statistics
University of Melbourne
Melbourne, Australia
qihanzou@gmail.com

Jie Li

School of Computing Technologies
RMIT University
Melbourne, Australia
hey.jiel@rmit.edu.com

Ke Deng

School of Computing Technologies
RMIT University
Melbourne, Australia
ke.deng@rmit.edu.au

Mark Sanderson

School of Computing Technologies
RMIT University
Melbourne, Australia
mark.sanderson@rmit.edu.au

Yongli Ren

School of Computing Technologies
RMIT University
MELBOURNE, VIC, Australia
yongli.ren@rmit.edu.au

Abstract

Recent progress in quantum computing has advanced research in quantum-assisted information retrieval and recommender systems, especially for feature selection via Quadratic Unconstrained Binary Optimization (QUBO). However, while existing work primarily focuses on effectiveness and efficiency, it often neglects the inherent noise and probabilistic nature of quantum hardware. In this paper, we propose a method based on Extreme Value Theory (EVT) to estimate the number of quantum executions (shots) needed to reliably obtain high-quality solutions—comparable to or better than classical baselines. Experiments on both simulators and two physical quantum devices demonstrate that our method effectively estimates the number of required runs to obtain satisfactory solutions on two widely used benchmark datasets.

CCS Concepts

• Mathematics of computing → Probability and statistics; • Information systems → Recommender systems; • Computer systems organization → Quantum computing.

Keywords

Quantum computing, Recommender systems, Feature selection, Extreme value theory

ACM Reference Format:

Jiayang Niu, Qihan Zou, Jie Li, Ke Deng, Mark Sanderson, and Yongli Ren. 2025. Estimating Quantum Execution Requirements for Feature Selection in Recommender Systems Using Extreme Value Theory. In *Proceedings of the Nineteenth ACM Conference on Recommender Systems (RecSys '25), September 22–26, 2025, Prague, Czech Republic*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3705328.3748029>



This work is licensed under a Creative Commons Attribution 4.0 International License.
RecSys '25, Prague, Czech Republic

© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-1364-4/25/09
<https://doi.org/10.1145/3705328.3748029>

1 Introduction

Recent advancements in quantum technology, along with growing media attention toward commercial quantum computers, suggest that qubit-based hardware may not only complement but in certain domains may outperform traditional transistor-based computing systems that have dominated for decades [1, 16]. Researchers have begun to explore how these quantum techniques can benefit Information Retrieval and Recommender Systems (RecSys). For instance, Nembrini et al. [12] embedded several classical feature selection strategies into Quadratic Unconstrained Binary Optimization (QUBO) formulations, while Niu et al. [20] applied counterfactual analysis to assess individual feature contributions to model performance and integrated these insights into their QUBO model. However, these studies primarily focus on the effectiveness and the efficiency when using quantum annealers (e.g., D-Wave), leaving two key *research gaps*: (1) Because quantum measurements are inherently probabilistic, it is necessary to run many shots¹, so as to determine the most likely or optimal results. For a fixed problem (e.g., feature selection in RecSys), it remains unclear how many shots are required to obtain high-quality results. (2) Compared to quantum annealers, gate-based quantum computers are more powerful and versatile, as they support a broader class of quantum algorithms beyond optimisation. Thus, exploring their use in this domain holds significant potential.

In this paper, we aim to answer the research question: *How to estimate the number of required shots for gate-based quantum computers to achieve good results in feature selection for recommendation tasks?* Specifically, we propose Quantum-EVT, a practical and theoretically grounded framework based on Extreme Value Theory (EVT) [17] for estimating the required number of quantum executions (shots), which provides confidence-based estimates that predict when quantum solvers are likely to outperform classical baseline model in terms of energy optimisation. Following existing work [12, 19, 20], we frame the feature selection problem in recommendation as a QUBO formulation for a click-through-rate prediction task and solve it using the Quantum Approximate Optimisation Algorithm (QAOA) [11], which is designed to solve

¹A shot refers to a single execution of a quantum program, including measurement. Multiple shots are used to sample from the output distribution.

combinatorial problems on a gate-based quantum device. However, there are two main challenges: (1) Deploying EVT requires several key statistical parameters, which must be estimated from many quantum executions to obtain some extreme value samples. While more samples generally lead to better results, there is a trade-off with computational cost. Thus, determining the minimum sufficient number of extreme samples becomes essential. (2) The quantum energy values derived from QUBO problems are inherently discrete, while EVT assumes a continuous distribution. Therefore, we must adapt the discrete quantum outputs to be compatible with EVT. To overcome these challenges, (1) we propose an algorithm that statistically determines the minimal number of extreme samples needed for stable EVT parameter estimation; (2) we introduce a technique to approximate a continuous distribution by adding uniform noise to the discrete quantum energy values.

The main contributions of this work are: (i) An EVT-based framework for estimating the required number of quantum executions, providing confidence-based shot estimates that predict when quantum solvers outperform classical simulated annealing (SA). (ii) The first application of gate-based quantum computers to solve the feature selection problem in recommender systems. (iii) A comprehensive evaluation of EVT-based predictions, demonstrating that the estimated number of runs is both accurate and robust under realistic settings, which also validates EVT as a robust tool for modelling quantum optimisation behaviour.

2 RELATED WORKS

Recent studies have explored the use of quantum optimization, particularly Quadratic Unconstrained Binary Optimization (QUBO), for feature and instance selection in information retrieval [21] (IR) and recommender systems (RecSys). CQFS [19], MIQUBO, CoQUBO, QUBO-Boosting [12], and PDQUBO [20] all formulate the feature selection task as a QUBO problem by designing different strategies for constructing the coefficient matrix Q . These strategies range from leveraging collaborative signals [19], to statistical measures such as mutual information and correlation [12], to performance-oriented counterfactual analysis [20]. These works also emphasize that quantum annealers, particularly D-Wave systems, are well-suited to solving QUBO problems due to their hardware design. However, PDQUBO additionally notes that annealer performance deteriorates as problem size increases, revealing stability concerns. Zaborniak et al. [26] also quantitatively benchmark the Hamiltonian noise on different generations of D-Wave hardware, showing significantly higher noise levels in newer systems (e.g., Advantage_system1.1) compared to earlier models.

Unlike quantum annealers, gate-based quantum computers are universal devices capable of running a broader class of quantum algorithms. The Quantum Approximate Optimization Algorithm (QAOA) [11] is a prominent candidate for solving QUBO problems on such devices. QAOA leverages parameterized variational circuits to approximate optimal solutions, enabling hybrid quantum-classical optimization pipelines. To complement the empirical use of QAOA, we adopt *Extreme Value Theory* (EVT) [5, 13] as a principled statistical framework to model the distribution of low-energy outcomes obtained from repeated quantum executions. EVT naturally fits the structure of quantum optimization, allowing us to

estimate how many repetitions are required to reach a target solution quality (e.g., matching classical baselines) under a specified confidence level.

3 PRELIMINARIES

Quadratic Unconstrained Binary Optimization (QUBO) is a mathematical framework for representing binary optimization problems. It optimizes a binary vector $\mathbf{x} \in \{0, 1\}^n$ (with the i -th entry x_i denoting the i -th feature) over a symmetric matrix $Q \in \mathbb{R}^{n \times n}$:

$$\min_{\mathbf{x}} Y = \mathbf{x}^T Q \mathbf{x} + \left(\sum_{i=1}^n x_i - k \right)^2, \quad \mathbf{x} \in \{0, 1\}^n \quad (1)$$

where Y is the ground state energy in quantum solutions, and the penalty term $(\sum_i x_i - k)^2$ enforces a cardinality constraint, ensuring that exactly k features are selected. This constraint-augmented QUBO formulation is particularly well-suited for tasks such as feature selection [12], where selecting a fixed-size subset of features is required. The resulting energy Y from QUBO optimization serves as a proxy for the quality of the selected subset—the lower the energy, the better the expected recommendation performance.

Quantum Approximate Optimization Algorithm (QAOA) [11] is a hybrid quantum-classical method designed for solving combinatorial optimization problems on gate-based quantum devices. To apply QAOA to a QUBO problem, binary variables $x_i \in \{0, 1\}$ are first mapped to spin variables $z_i \in \{-1, 1\}$ via $x_i = \frac{1+z_i}{2}$. Substituting this into the QUBO formulation yields an equivalent objective expressed entirely in terms of z_i :

$$\begin{aligned} \min_{\mathbf{x}} Y &= \frac{1}{4} \sum_{i,j} Q_{ij} (1+z_i)(1+z_j) + \frac{1}{4} \left(\sum_{i=1}^n (1+z_i) - 2k \right)^2, \\ &= \frac{1}{2} \sum_{i=1}^n \left(\sum_{j=1}^n Q_{ij} + n - 2k \right) z_i + \frac{1}{4} \sum_{i,j=1}^n (Q_{ij} + 1) z_i z_j \end{aligned} \quad (2)$$

This transformation converts the QUBO objective into an Ising Hamiltonian composed of single-qubit (Z_i) and two-qubit ($Z_i Z_j$) terms. QAOA constructs a parameterized quantum state by alternately applying two unitaries: a cost operator $U_C(\gamma) = e^{-i\gamma H_C}$, where $H_C = \sum_i h_i Z_i + \sum_{i < j} J_{ij} Z_i Z_j$ encodes the problem structure, and a mixer operator $U_M(\beta) = e^{-i\beta H_M}$, where $H_M = \sum_i X_i$ enables state exploration. These operations are applied in p layers to an initial uniform superposition $|\psi_0\rangle = H^{\otimes n} X^{\otimes n} |0\rangle^{\otimes n}$. The variational parameters (γ, β) are optimized classically to minimize the expected energy $\langle \psi(\gamma, \beta) | H_C | \psi(\gamma, \beta) \rangle$. Low-energy measurement outcomes correspond to better approximate solutions to the QUBO problem.

PDQUBO [20] formulates feature selection as a QUBO problem by encoding the impact of individual and joint feature removals on model performance into the matrix Q in Eq. 1. Specifically, the diagonal and off-diagonal elements of Q are computed as: $Q_{ii} = G(\mathcal{F}) - G(\mathcal{F}_{\text{mask}}^i)$ when $i = j$; otherwise $Q_{ij} = G(\mathcal{F}) - G(\mathcal{F}_{\text{mask}}^{ij})$,

where $G(\cdot)$ is the performance metric (e.g., Area Under the Curve - AUC) of a trained recommendation model. $\mathcal{F}_{\text{mask}}^i$ and $\mathcal{F}_{\text{mask}}^{ij}$ denote feature subsets with one or two features removed. Thus, Q reflects both individual feature relevance and pairwise interactions.

4 Quantum-EVT (Q-EVT)

We propose Q-EVT (as shown in Figure 1), a framework for estimating the number of quantum executions (shots) required for a given task. It formulates feature selection in recommendation as a QUBO problem, solved using QAOA to obtain extreme value samples for applying Extreme Value Theory (EVT). Q-EVT outputs confidence-based estimates predicting when quantum solvers are likely to outperform classical models in energy optimisation.

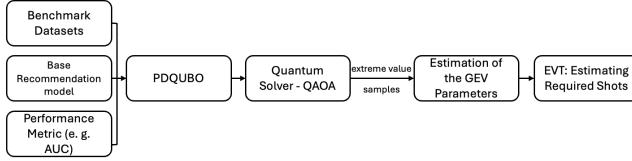


Figure 1: The Overview of Q-EVT

4.1 Extreme Value Theory (EVT)

QAOA yields probabilistic outputs, where each execution (*shot*) returns only one candidate solution. Due to noise and inherent randomness, optimal solutions are obtained only with limited probability, motivating the use of repeated shots. Extreme Value Theory (EVT) [17, 18, 23, 24] offers a natural framework to analyze the minimum energy values obtained across these repeated executions. By modeling the distribution of such extreme outcomes, EVT enables principled estimation of quantum resources required to achieve desired solution quality.

Assume energy samples $\{y_1, y_2, \dots, y_n\}$ from repeated quantum executions are independent and identically distributed (i.i.d.) from an unknown distribution F . EVT primarily analyzes the distribution of block maxima [13, 15]:

$$M_n = \max(y_1, y_2, \dots, y_n). \quad (3)$$

Under certain conditions, the normalized block maximum M_n converges to a Generalized Extreme Value (GEV) distribution [15]:

$$\mathbb{P}\left(\frac{M_n - \mu}{\sigma} \leq z\right) \rightarrow G(z) = \begin{cases} \exp\left(-\left[1 + \xi \frac{z - \mu}{\sigma}\right]^{-1/\xi}\right), & \xi \neq 0, \\ \exp\left(-\exp\left(-\frac{z - \mu}{\sigma}\right)\right), & \xi = 0, \end{cases} \quad (4)$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter, and $\xi \in \mathbb{R}$ is the shape parameter. While the original formulation focuses on maxima, analyzing minima can be achieved by negating observations ($-y_i$). However, accurately applying EVT demands reliable estimation of GEV distribution parameters.

4.2 Estimating Required Extreme Samples

The estimation of the GEV parameters (μ , σ , and ξ) requires a sufficient number of extreme value samples. While larger sample sizes improve the stability of parameter estimates, practical constraints—such as the cost and latency of quantum executions necessitate a careful trade-off. Thus, identifying the minimal number of samples required for reliable estimation becomes essential. To address this, we follow the previous work by [4] (that provided an estimation method for minimum samples with return level as

the objective with univariate Shapiro-Wilk normality test) and the idea of the bootstrap method [7, 8] to design a statistical procedure that estimates the smallest sufficient sample size n_{estimate} . Starting with a large set of extreme values Y_{sim} (Each value in Y_{sim} is a block-wise extreme like M_n , collected from repeated quantum executions) obtained from quantum simulations, we compute a stable reference GEV parameter set $\theta_{\text{sim}} = (\mu_{\text{sim}}, \sigma_{\text{sim}}, \xi_{\text{sim}})$. Then, for a range of candidate sample sizes $n \in [n_{\text{min}}, N_{\text{max}}]$, we repeatedly draw random subsets from Y_{sim} , estimate GEV parameters for each subset, and perform two statistical tests on the resulting estimates: Hotelling's T^2 test [2] to check for consistency with the reference mean and the multivariate Shapiro-Wilk test [25] to assess normality.

The average p -values from these tests indicate the stability of parameter estimation at each n . The smallest n for which both tests yield p -values exceeding 0.05 is selected as n_{estimate} . Note: N_{max} should be chosen conservatively. If the subset size approaches the total pool N , the samples become nearly identical due to the limited size of discrete quantum outcomes. This causes the multivariate Shapiro-Wilk test to fail, as the sampled distributions collapse to a point. While assuming continuous distributions could mitigate this, such an assumption is generally unjustified given the discrete nature of quantum outputs. The complete procedure is summarized in Algorithm 1.

4.3 Estimation of the Required Shots

Quantum energy samples from QAOA are inherently discrete due to the binary nature of solution vectors, whereas EVT assumes continuous distributions. To address this, we add uniform noise to each energy sample: $\tilde{y}_i = y_i + \epsilon_i$, where $\epsilon_i \sim \mathcal{U}(-\delta/2, \delta/2)$ and δ is the smallest nonzero difference between sorted unique energy values [5, 9, 10, 15]. This smoothing step allows EVT to more accurately model the distribution of extreme values.

Given a reference energy value y_{ideal} obtained from classical Simulated Annealing (SA), we estimate how many quantum executions (each with s shots) are required to achieve, with confidence level α , at least one result no worse than y_{ideal} . Letting $p = \mathbb{P}(y \leq y_{\text{ideal}})$ denote the probability that a single quantum execution yields a satisfactory result (as determined via the fitted GEV distribution), the required number of executions n_{EVT} satisfies:

$$1 - (1 - p)^{n_{\text{EVT}}} \geq \alpha \quad \Rightarrow \quad n_{\text{EVT}} \geq \frac{\log(1 - \alpha)}{\log(1 - p)}.$$

The total number of required quantum shots is then simply $n_{\text{EVT}} \times s$. s denotes the number of shots executed to compute a single extreme point y_i .

5 EXPERIMENT

5.1 Experiment Setup

5.1.1 *Quantum Devices and Algorithm Configuration.* Experiments are conducted on two gate-based quantum devices via AWS Braket: *Ankaa-3* (superconducting) and *Forte Enterprise 1* (ion-trap), and a simulator that is also offered by AWS Braket to be consistent with

Algorithm 1: Estimate Required Extreme Samples

Input : Y_{sim} – sequence of simulated extreme values
 n_{\min}, N_{\max} – range of sample sizes to test
 I, J – number of repetitions per estimation
 $\theta_{\text{sim}} = (\mu_{\text{sim}}, \sigma_{\text{sim}}, \xi_{\text{sim}})$ – reference GEV parameters
Output: n_{estimate} – estimated number of extreme samples required

```

1 for  $n = n_{\min}$  to  $N_{\max}$  do
2   for  $j = 1$  to  $J$  do
3     for  $i = 1$  to  $I$  do
4       Draw  $n$  samples  $Y_{ijn}$  from  $Y_{\text{sim}}$  with replacement;
5       Estimate GEV parameters from  $Y_{ijn}$ :
6          $\hat{\theta}_{ijn} = (\hat{\mu}_{ijn}, \hat{\sigma}_{ijn}, \hat{\xi}_{ijn})$ ;
7       Let  $\hat{\theta}_{jn}$  denote the collection of estimates for
8          $i = 1, \dots, I$ ;
9       Perform Hotelling's  $T^2$  test for the null hypothesis
10       $H_0 : E(\hat{\theta}_{jn}) = \theta_{\text{sim}}$ ,
11      and obtain the p-value  $p_{jn,ht2}$ ;
12      Perform the multivariate Shapiro-Wilk normality
13      test on  $\hat{\theta}_{jn}$  and obtain the p-value  $p_{jn,mst}$ ;
14      Compute the average p-values:
15       $\bar{p}_{n,ht2} = E(p_{jn,ht2})$  and  $\bar{p}_{n,mst} = E(p_{jn,mst})$ 
16      over  $j = 1, \dots, J$ ;
17      Fit regression lines to the pairs  $(n, \bar{p}_{n,ht2})$  and  $(n, \bar{p}_{n,mst})$ ;
18      Determine the smallest  $n$  such that both  $\bar{p}_{n,ht2}$  and  $\bar{p}_{n,mst}$ 
19      exceed 0.05 based on the intersection of regression lines
20      and  $p = 0.05$ , and let these be  $n_{ht2}$  and  $n_{mst}$  respectively;
21      if  $n_{ht2} < n_{mst}$  then
22         $n_{\text{estimate}} \leftarrow n_{mst}$  // Hotelling's  $T^2$  test requires
23        normality
24      else
25         $n_{\text{estimate}} \leftarrow n_{ht2}$ 
26      return  $n_{\text{estimate}}$ ;

```

these platforms². The simulator performs ideal (noise-free) quantum circuit emulation using the native gate sets of each corresponding quantum device. We solve QUBO instances using the Quantum Approximate Optimization Algorithm (QAOA) with a circuit depth of $p = 3$, optimized using COBYLA [22]. All QAOA parameters are first tuned on a simulator, from which the best-performing set is transferred to real quantum devices. In real device, all experiments use 2000 shots. Additional platform-specific implementation details are available in the Braket SDK documentation³.

5.1.2 Datasets and Base model. We use two standard click-through-rate (CTR) prediction datasets: Avazu and Criteo [3, 6]. To match quantum hardware limitations, we construct feature subsets of sizes

²AWS Quantum Device

³Braket SDK Docs

10, 13, 15, and 18. Feature selection is guided by XGBoost importance scores: the top 3 and bottom 2 features are retained, and the remaining are sampled from importance-ranked groups to ensure diversity. QUBO matrices Q are derived from a base recommendation mode (DeepFM [14]) following the PDQUBO framework. We also impose cardinality constraints on the number of selected features (as shown in Eq. 1), set to 8, 11, 12, and 14 for each subset size, respectively. Our code is available at https://github.com/jiayangniu/Recsys25_QEVT

Table 1: Estimated executions n_{EVT} needed for quantum simulators and Quantum Processing Unit (QPU) to match or surpass classical SA at 95% and 90% confidence, across various feature selection scales

Dataset	Avazu		Criteo			
	simulator	Ankaa-3	Forte	Ankaa-3	Forte	
Scale	shots	95% /90% >= SA				
		500	7/6	13/10	6/5	10/8
		1000	5/4	9/7	4/3	6/5
10	2000	500	4/3	5/4	5/4	4/3
		1000	23/16	24/16	15/12	24/19
		2000	13/11	16/13	7/6	13/11
13	2000	500	8/6	9/7	5/4	7/6
		1000	62/47	103/80	62/48	70/54
		2000	30/25	71/54	34/26	36/28
15	2000	500	19/15	46/35	22/17	31/24
		1000	Ankaa-3	Forte	Ankaa-3	Forte
		2000	16/13	25/19	23/17	31/24
QPU	500	13/11	27/21	68/52	51/39	97/75
		1000	8/6	9/7	5/4	7/6
		2000	62/47	103/80	62/48	70/54
10	500	30/25	71/54	34/26	36/28	
		1000	19/15	46/35	22/17	31/24
		2000	76/58	111/87	114/88	209/161

5.2 Experiment Results

5.2.1 EVT-Based Shot Estimation. We conduct experiments on the Avazu and Criteo datasets using feature subset sizes of [10, 13, 15]. We compare against classical simulated annealing (SA) since prior QUBO studies have commonly used SA as a performance baseline for quantum optimization [12, 20]. For each configuration, we run quantum simulations under shot settings of [500, 1000, 2000] to estimate the number of runs required to reach classical SA performance using EVT. On real quantum hardware (Ankaa-3 and Forte), we execute only the 500-shot setting for selected scales. The required number of extreme samples to fit the GEV distribution is estimated using the procedure described in Section 4.2, which suggests a threshold of 70–80 samples for stable estimation. To ensure robustness, we conservatively collect 200 extreme samples per configuration. Each sample corresponds to the minimum energy obtained from a single quantum run. These values are then fitted to a GEV model, which is used to estimate the total number of quantum executions needed to exceed SA performance with 95% and 90% confidence levels.

Table 1 presents the estimated number of runs (blocks of shots) required for quantum simulators and real quantum devices to match or exceed classical simulated annealing (SA) performance at 95% and 90% confidence levels across different feature selection scales. A clear trend emerges: as the feature selection scale increases, more

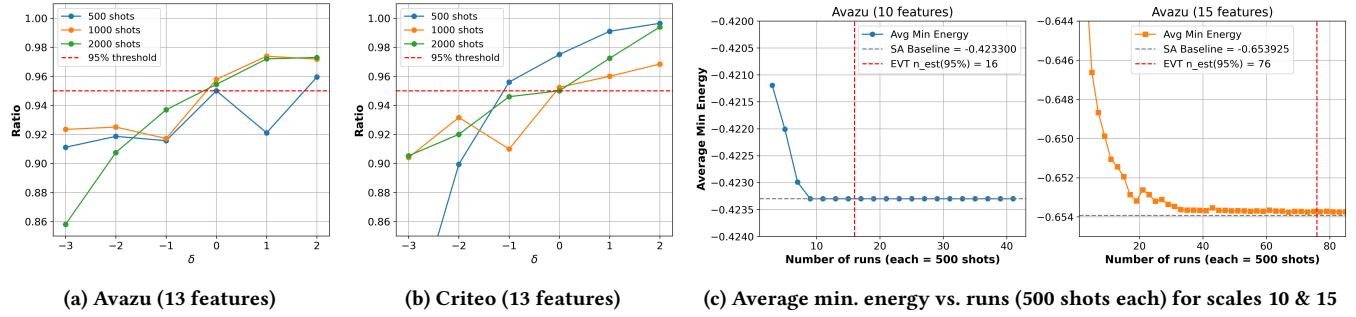


Figure 2: (a) and (b) Evaluation on simulators, and (c) Evaluation and trend on physical quantum processor Ankaa-3. Note the red dashed line denotes the 95% confidence target.

runs are required. For instance, on the Avazu dataset using the Ankaa-3 simulator with 500 shots per run, the required number of runs under 95% confidence grows from 7 at scale 10 to 62 at scale 15. Similarly, the Forte simulator requires 13 and 103 runs, respectively, over the same scales. Additionally, increasing the number of shots per run reduces the required number of runs for a fixed scale. For example, at scale 13 on Avazu, the Ankaa-3 simulator requires 23, 13, and 8 runs when using 500, 1000, and 2000 shots, respectively. Since each configuration was independently sampled, this monotonic reduction supports the consistency and reliability of EVT-based estimation. Finally, real quantum devices require more runs than simulators, likely due to noise. On Criteo at scale 15 with 500 shots, the Ankaa-3 simulator requires 62 runs for 95% confidence, while the physical Ankaa-3 device requires 114 runs. This gap reflects real hardware imperfections and reinforces the need for robust estimation in practical quantum optimization.

5.2.2 Evaluation on Simulator. To test whether our EVT model accurately predicts the number of quantum executions required to match or exceed classical SA performance, we conducted 1,000 independent QAOA runs on the Avazu and Criteo datasets at a fixed feature scale of 13. For each dataset, we used three per-run shot budgets $s \in \{500, 1000, 2000\}$ and measured, for each total-shot count ($n_{\text{EVT}} + \delta$) $\times s$, the fraction of runs whose minimal energy fell at or below the SA baseline, which is defined as our empirical “ratio”. Then, we plotted these ratios against δ from the corresponding Q-EVT run count n_{EVT} in Figures 2a and 2b. Specifically, an accurate prediction implies that the curve crosses the horizontal dashed line precisely at $\delta = 0$. Indeed, for both the 500-shot and 2000-shot settings, the empirical curves intersect at $\delta \approx 0$, confirming that our EVT estimates pinpoint the required number of repetitions with high precision. While the 1000-shot configuration on Criteo exhibits slightly more scatter, its crossing remains very close to $\delta = 0$, demonstrating that the method remains robust even under realistic device noise and sampling variability.

5.2.3 Evaluation on Physical Quantum Processors. We evaluate this by demonstrating the trend of optimisation results while increasing the number of runs on a real quantum processor. Using the Ankaa-3 superconducting QPU, we solve the Avazu feature-selection problem at two fixed scales (13 and 15 features). For each scale, we sweep the per-run shot budget s from 500 to 20000 in steps of 1000. At each

s, we perform 20 independent QAOA executions, record the minimum energy observed in each execution, and plot the average of these 20 minima. Figure 2c shows that, for both scales, the average best-observed energy steadily approaches the classical SA baseline (dashed line) as s increases. In particular, small shot budgets (below 5000) yield noticeably higher energies, while beyond $\sim 10\,000$ shots the averaged minimum energy nearly converges to the SA reference. These results demonstrate that, even on noisy gate-based quantum hardware, simply allocating more shots per run can substantially improve the probability of finding high-quality solutions.

5.2.4 Limitations and Breakdown Scenarios. While our EVT-based framework provides accurate shot estimates across a variety of settings, we observe two scenarios where the approach may begin to break down. First, when the quantum algorithm consistently yields identical optimal outputs within a small number of shots, the lack of output variability limits the statistical basis required for reliable EVT modeling. Second, when quantum outputs remain significantly suboptimal even after many executions, the estimated shot count may become unstable or overestimated due to poor alignment with the classical solution landscape.

6 Conclusion

Our results demonstrate that Extreme Value Theory (EVT) provides an effective framework for estimating the number of quantum executions (runs) required to match or exceed classical simulated annealing (SA) performance. Across both quantum simulators and real gate-based devices, the predictions of Q-EVT align with empirical outcomes at various feature selection scales. This allows for robust, confidence-based shot estimation without requiring exhaustive empirical tuning.

Acknowledgments

This research is supported by the ARC Discovery Project (DP210100743) and RACE (RMIT Advanced Cloud Ecosystem). We would like to express our sincere gratitude to RACE for providing access to quantum computing hardware and computational infrastructure, which was essential for conducting our experiments and advancing this research.

References

[1] Google Quantum AI et al. 2024. Quantum error correction below the surface code threshold. *Nature* 638, 8052 (2024), 920.

[2] Theodore Wilbur Anderson, Theodore Wilbur Anderson, Theodore Wilbur Anderson, Theodore Wilbur Anderson, and Etats-Unis Mathématicien. 1958. *An introduction to multivariate statistical analysis*. Vol. 2. Wiley New York.

[3] Avazu Inc. 2015. Avazu CTR Prediction. <https://www.kaggle.com/c/avazu-ctr-prediction> Accessed: 2025-07-15.

[4] Yuzhi Cai and Dominic Hames. 2010. Minimum sample size determination for generalized extreme value distribution. *Communications in Statistics—Simulation and Computation®* 40, 1 (2010), 87–98.

[5] Stuart Coles, Joanna Bawa, Lesley Trenner, and Pat Dorazio. 2001. *An introduction to statistical modeling of extreme values*. Vol. 208. Springer.

[6] Criteo Labs. 2014. Criteo Labs Display Advertising Challenge. <https://www.kaggle.com/c/criteo-display-ad-challenge> Accessed: 2025-07-15.

[7] Anthony Christopher Davison and David Victor Hinkley. 1997. *Bootstrap methods and their application*. Number 1. Cambridge university press.

[8] Bradley Efron and Robert J Tibshirani. 1994. *An introduction to the bootstrap*. Chapman and Hall/CRC.

[9] Davide Faranda, Jorge Milhazes Freitas, Valerio Lucarini, Giorgio Turchetti, and Sandro Vaienti. 2013. Extreme value statistics for dynamical systems with noise. *Nonlinearity* 26, 9 (Aug. 2013), 2597–2622. doi:10.1088/0951-7715/26/9/2597

[10] Davide Faranda and Sandro Vaienti. 2014. Extreme value laws for dynamical systems under observational noise. *Physica D: Nonlinear Phenomena* 280 (2014), 86–94.

[11] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. 2014. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028* (2014).

[12] Maurizio Ferrari Dacrema, Fabio Moroni, Riccardo Nembrini, Nicola Ferro, Guglielmo Fagioli, and Paolo Cremonesi. 2022. Towards feature selection for ranking and classification exploiting quantum annealers. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2814–2824.

[13] M Ivette Gomes and Armelle Guillou. 2015. Extreme value theory and statistics of univariate extremes: a review. *International statistical review* 83, 2 (2015), 263–292.

[14] Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. 2017. DeepFM: a factorization-machine based neural network for CTR prediction. *arXiv preprint arXiv:1703.04247* (2017).

[15] Laurens Haan and Ana Ferreira. 2006. *Extreme value theory: an introduction*. Vol. 3. Springer.

[16] Andrew D King, Alberto Nocera, Marek M Rams, Jacek Dziarmaga, Roeland Wiersema, William Bernoudy, Jack Raymond, Nitin Kaushal, Niclas Heinsdorf, Richard Harris, et al. 2025. Beyond-classical computation in quantum simulation. *Science* (2025), eado6285.

[17] Samuel Kotz and Saralees Nadarajah. 2000. *Extreme value distributions: theory and applications*. world scientific.

[18] Marie Kratz. 2019. Introduction to extreme value theory: Applications to risk analysis and management. *2017 MATRIX Annals* (2019), 591–636.

[19] Riccardo Nembrini, Maurizio Ferrari Dacrema, and Paolo Cremonesi. 2021. Feature selection for recommender systems with quantum computing. *Entropy* 23, 8 (2021), 970.

[20] Jiayang Niu, Jie Li, Ke Deng, Mark Sanderson, Nicola Ferro, and Yongli Ren. 2025. Performance-Driven QUBO for Recommender Systems on Quantum Annealers. *arXiv:2410.15272 [cs.IR]* <https://arxiv.org/abs/2410.15272>

[21] Andrea Pasin, Washington Cunha, Marcos André Gonçalves, and Nicola Ferro. 2024. 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*. 205–214.

[22] Michael JD Powell. 1994. *A direct search optimization method that models the objective and constraint functions by linear interpolation*. Springer.

[23] Richard L Smith. 1990. Extreme value theory. *Handbook of applicable mathematics* 7, 437–471 (1990), 18.

[24] Richard L Smith. 1994. *Extreme values*. Chapman & Hall London, UK.

[25] José A Villasenor Alva and Elizabeth González Estrada. 2009. A generalization of Shapiro–Wilk’s test for multivariate normality. *Communications in Statistics—Theory and Methods* 38, 11 (2009), 1870–1883.

[26] Tristan Zaborniak and Rogério de Sousa. 2021. Benchmarking Hamiltonian noise in the D-Wave quantum annealer. *IEEE Transactions on Quantum Engineering* 2 (2021), 1–6.