URL Classification using Quantum Classifier

***Abstract***

*Quantum computing provides an intriguing scenario in which it may be able to provide certain enhancements and additions to a traditional network as it is being trained. This approach is particularly common in the present Noisy Intermediate-Scale Quantum age when we can test these hypotheses using ML frameworks like TensorFlow and libraries like Pennylane. This paper presents a proof-of-concept for the same, using a hybrid quantum-classical model to solve a text classification problem on the extracted data from URLs. This model takes a clustered text with a unique set of texts that a particular URL has, we encode this textual data via Universal Sentence Encoder and running a Softmax function, via a quantum circuit we get the density of words and display the output as a single text which is the most probable representation of the whole cluster.*

**Introduction**

When it comes to current natural science, the quantum mechanics theory (QM) is one of the most recognized and complex theories that put a lot of things together. Since QM, there has been a world-shaking revolution in physics. was born, it has been regarded as a significant component of theoretical physics and has demonstrated its use in describing experimental results Furthermore, Some scientists feel that QM is the final principle of physics, if not all natural science. As a result, the number of researchers has grown. the study of QM in other domains of research, and has profoundly influenced practically every field of natural science and technology, such as quantum mechanics computation.

The quantum computation concept has also influenced many scientific studies in computer science, notably computational modeling, cryptography theory, and information theory. Some researchers have used quantum computation principles and technology to improve studies on Machine Learning (ML) (Ameur et al., 2006; Ameur et al., 2007; Chen et al., 2008; Gambs, 2008; Horn and Gottlieb, 2001; Nasios and Bors, 2007), a field that studies theories and constructions of systems that can learn from data, of which classification is a typical task. Thus, in order to demonstrate the viability of using the QM model to machine learning, we sought to construct a computational model based on quantum computing theory to handle categorization problems.

In this paper, we describe a way of treating the classifier as a physical system accessible to QM and the entire classification process as the evolutionary process of a closed quantum system. A unitary operator can characterize the development of a quantum system, according to QM.

**1) Classification of data in QC**

As mentioned in [2] Classification is a prominent problem in the supervised learning domain that transfers provided input data (x) to discrete target output (y) via a function approximation f(y) = f(x). The primary goal of categorization is to create an accurate prediction model. Classification issues can be divided into two categories: Binary classification refers to classification with two classes. Multi-class classification refers to classification having more than two classes (for instance image and digit classification).

Classification problem can be represented with classical ML domain as C = {𝑐1, 𝑐2, …, 𝑐n} where C is target labels and a set of data in training phase with as 𝐷𝑛= {(𝑥1, 𝑦1), …, (𝑥i, 𝑦𝑖 ), …, (𝑥n, 𝑦𝑛)} where xi is some of the features (n) on the properties of an order of data point (i) and 𝑦𝑖 is the corresponding of that data point. In the case of binary classification 𝑦𝑖∈ {𝑐1, 𝑐2 } and xi ∈ 𝑅 𝑑 . in the case of multi-class classification 𝑦𝑖 ∈ {𝑐1, ….., 𝑐n } where xi ∈ 𝑅 𝑑 and d is real-valued attributes. To can describe classification problems with the QML domain, we should convert classical data to quantum data then we can be represented quantum data in training data as 𝐷𝑛 = {( |𝜓1 ⟩, 𝑦1 ), …, ( |𝜓𝑖 ⟩, 𝑦𝑖 ),…, ( |𝜓𝑛 ⟩, 𝑦𝑛)} where |𝜓𝑖 ⟩ is the order (i) of the quantum state of 𝐷𝑛, |𝜓𝑖 ⟩ ∈ ∁ 2 𝑑 and 𝑦𝑖∈ {c1, c2 } [10], [11] in the case of binary classification.

To work with the quantum data, we need to convert our classical data into quantum data.

**2) Methods to Encode Classical Data into Quantum Data**

Quantum computers are designed to process data in the quantum state. NISQ (Noisy Intermediate Scale Quantum) devices have a restricted number of qubits that are stable for a short length of time. Loading conventional data into the state of the Qubits is the initial stage in Quantum Machine Learning. This procedure, also known as quantum data encoding or embedding, is a key stage in the production of quantum states. The importance of classical data encoding in quantum computation is crucial to the overall design and performance of the Quantum Machine Learning algorithm (QML).

To use modern NISQ devices, the representation must be small and employ only a few qubits and quantum gates. Qubits not only decay quickly, but quantum gates are also error-prone, restricting the number of operations required to establish the quantum state, which must be modest.

**Three phases are included in Quantum Machine Learning.**

* Encoding: the process of loading classical data into a quantum state.
* Processing: The embedded input, which will be a variational circuit or a quantum routine, is processed by the Quantum device at this point.
* Measurement: This stage measures the predicted result, which subsequently forms the forecast for QML.



We are using, Google released a pre-trained version of the [Universal Sentence Encoder](https://arxiv.org/abs/1803.11175) (USE) model. As mentioned in [15] We describe methods for encoding words into embedding vectors with the goal of transferring learning to other NLP tasks. The models are efficient and produce reliable results on a variety of transfer tasks. Two encoding model versions allow for trade-offs between accuracy and computation resources. We explore and report on the link between model complexity, resource consumption, the availability of transfer task training data, and task performance for both types. Baselines that employ word-level transfer learning via pre-trained word embeddings are compared to baselines that do not use any transfer learning. We discover that transfer learning with sentence embeddings outperforms word-level transfer. We see an unexpectedly strong performance with minimum quantities of supervised training data for a transfer task using transfer learning via phrase embeddings.



**3) Methodology**

****

**4) Dataset Description**

1. ***Dataset Discussion***

1.1 Sample of Tor Hidden Services

| **Link:** <http://222222222xn2ozdb2mjnkjrvcopf5thb6la6yj24jvyjqrbohx5kccid.onion/>  **Title:** Best Financial Service - #1 shop to earn risk free money for anybody!  **ScreenShot:** |
| --- |

| **Link**: <http://2a2a2abbjsjcjwfuozip6idfxsxyowoi3ajqyehqzfqyxezhacur7oyd.onion/>  **Title:** Empire Market – Best darknet market | Credit Card Cloned Carding Hacking Drugs Dumps Paypal Hack Free Bitcoin Money Counterfeit Cash Buy Gun Gift Passport Visa Mastercard Amex Verified Trusted Bitcoins Escrow Top Hidden Wiki Onion Links Forum  **ScreenShot:** |
| --- |

1.2 *Data Collection*

The HTML pages of Tor links is scrapped down with the help of a crawler, The words are scrapped form HTML pages and cleaned by following processes:

* Removal of Stopwords.
* Large N-Gram Cleaning.
* Lemmatization and Stemming.
* Removal of Punctuation.
* Converting all the Capital letters to lowercase.

1.3 *Dark Web Crawler*

??? Crawler Made (Dark Web crawler)

Period (3 - 4 Months)

Size = 44,000 URLs

1.4 *Period of The Dataset*

??? 3-4 Months1.5 *Sample Dataset*

| Id | URLs | Cleaned\_Data | Title |
| --- | --- | --- | --- |
| 1 | crawled\4p6i33oqj6wgvzgzczyqlueav3tz456rdu632xzyxbnhq4gpsriirtqd.onion/3CB6NF7FCFC02A2CFB2C8E7EFBAF8E9.html | visit view visit use experience people completely new way b c d e f g h j k l m n o p q r s t u v w x y z check united kingdom grad visit united kingdom new  More… | Peoples Drug Store - The Darkweb's Best Online Drug Supplier! - Buy cocaine, speed, xtc, mdma, heroin and more at peoples drug store, pay with Bitcoin |
| 2 | crawled\canxzwmfihdnn7bz.onion/A7045E8B1F9E04E6GF35DF8b96C29.html | chat accessible new shiny v3 address notbumpz34bgbz4yfdigxvd6vzwtxc3zpt5imukgl6bvip2nikdmdaad onionnotbumpz34bgbz4yfdigxvd6vzwtxc3zpt5imukgl6bvip2nikdmdaad onion chat group chat  More… | Ableonion |
| 3 | crawled\bepig5bcjdhtlwpgeh3w42hffftcqmg7b77vzu7ponty52kiey5ec4ad.onion\F7E89E9EDFFByAKxE8c3E99C6s.html | quality original cheap paymentkamagra4bitcoin buy cheap ship login register login register mg generic popular successful widely accept treatment clinical clean room produce high quality standard ensure safety effectiveness regularly report successful  More… | Kamagra For Bitcoin - Same quality as original viagra pills, cheap prices, Bitcoin payment |
| 4 | crawled\ar-ar.facebookcorewwwi.onion\6ADH89JBD10FB2A1B1084DABFF424.html | solidarity foundation alt solidarity foundation triangle right triangle right stripe triangle right triangle right medical need help s medical help rick adopt kidney future flore little s medical boss wren solidarity foundation solid bycovid19 solidarity  More… | ‪Belarus Solidarity Foundation‬ |
| 5 | crawled\cs-cz.facebookcorewwwi.onion\nF1C65E89L8EP1BRB4021FEF9.html | fan ocean cleanup ocean cleanup ocean cleanup ocean cleanup advanced na k alt se na se e mail k se se na se se fan ocean cleanup message fan na od cake triangle right na na na v z na triangle right od se na triangle  More… | Ilza Fan The Ocean Cleanup |
| 6 | crawled\cx-ph.facebookcorewwwi.onion\BDF9CA93A50CD09FY9C2AkF441DG.html | sa sa kung sa sa ug kung sa sa log account home sa account home page watch mobile accessibility sa account ug password profile ug ad information account ug privacy safe keep account secure unfriending block ug  More… | Announcements | Facebook Help Center | Facebook |
| 7 | crawled\zqktlwiuavvvqqt4ybvgvi7tyo4hjl5xgfuvpdf6otjiycgwqbym2qad.onion\YD68EB1A0aB34823CE1700JW.html | talk link hide link hide jump navigation jump search zqktlwiuavvvqqt4ybvgvi7tyo4hjl5xgfuvpdf6otjiycgwqbym2qad onion index title talk navigation menu page page discussion view source history page page discussion  More… | Talk:Link proposals - The Hidden Wiki |
| 8 | crawled\bg-bg.facebookcorewwwi.onion\0BDB8C9ECFDF1A8DA7-11FE007Bl.html | alt alt alt alt alt alt alt | Facebook Fundraising Campaigns Category... |
| 9 | crawled\ca-es.facebookcorewwwi.onion\A029BF3B3287BCEzAD29989F6F086.html | de la de tu en d d alt charge tu la en de data policy lo en de de para la las y en n lo en mi n de mi de mi n y y en las de n en hay persona n de y de n tu en tu tu de la para de  More… | Aspectos básicos de la privacidad de Facebook |
| 10 | crawled\da-dk.facebookcorewwwi.onion\ACFU9C\_-0AC61298614DD2A77B.html | local til log e mail din log en lister live history tour weekday wind southwest neighborhood guide riverside neighborhood guide old town neighborhood guide comfort midday school neighborhood guide health kick happy hour  More… | Local Guides |
| 11 | crawled\de-de.facebookcorewwwi.onion\0Aa03F61DBFE6i5D1960F867BCE.html | friendly dich e mail listen friendly upscale eats thrifty barbecue table study break pamper game day frozen family fun dog friendly eats bounty budget scene friendly messenger lite watch pay  More… | Veg-Friendly Delights |
| 12 | crawled\zqktlwiuavvvqqt4ybvgvi7tyo4hjl5xgfuvpdf6otjiycgwqbym2qad.onion\YB1D849901dB6F9d8AEE048BD1l.html | network computer hidden computer hidden jump navigation jump search content mac address mac os nickname mac os mac os address proxy mac os tor i2p  More… | Network Attributes of your computer - The Hidden Wiki |

1.5 *Sample Clean Text*

| *URL:* 4p6i33oqj6wgvzgzczyqlueav3tz456rdu632xzyxbnhq4gpsriirtqd.onion/3CB6NF7FCFC02A2CFB2C8E7EFBAF8E9.html  *Cleaned Text:* drug store s good drug supplier buy cocaine speed heroin drug store pay drug store number deep web drug vendor buy register login register drug store pride offer good quality competitive make effort come customer satisfaction choose category follow heroin cocaine ecstasy speed free tell shop earn purchase simply follow link ref original onion ref replace actual site earning directly wallet heroin heroin offer come direct importer middle man white light beige color great pride fact cut product whatsoever ensure source prefer offer high quality product repeat s difference heroin people know decide add information listing commonly find heroin grade h form heroin white powder easily water readily grade h tan granular product brown rock difference color result process grade commonly citrus instead water simple think like h water pure people sniff people prefer smoke h smoke usually adulterant burn register drug store pride offer good quality competitive make effort come customer satisfaction choose category follow heroin cocaine ecstasy speed free tell shop earn purchase simply follow link ref original onion ref replace actual site earning directly wallet heroin heroin offer come direct importer middle man white light beige color great pride fact cut product whatsoever ensure source prefer offer high quality product repeat s difference heroin people know decide add information listing commonly find heroin grade h form heroin white powder easily water readily grade h tan granular product brown rock difference color result process grade commonly citrus instead water simple think like h water pure people sniff people prefer smoke h smoke usually adulterant burn  *Title:* Peoples Drug Store - The Darkweb's Best Online Drug Supplier! - Buy cocaine, speed, xtc, mdma, heroin and more at peoples drug store, pay with Bitcoin |
| --- |

1.5 *Classical Data Classification*

Using the BeautifulSoup package, a scraper tool has been used to parse and scrape every webpage. The text contents of each website, including the text within the <title> tag, the text within the <meta> tags related to keywords, the text within the description tag, the text within heading tags because they may contain relatively important text, and the text within the text of the entire page with the exception of some tags, were all gathered. Pronouns were eliminated and stopwords was cleaned out from the gathered texts from the webpages by using spacy library. The words were lowercased and lemmatized. Then, we will have a set of keywords that fall under a specific category, and so we will compare those keywords to the text to determine which group has the most matching keywords, to get the keywords inside the cleaned text we use KeywordProcess from a package flashtext on PyPI. To transform the data into features we use TfidfVectorizer to transform it into vector, lastly, to evaluate the accuracy of several models, including RandomForestClassifier, LinearSVC, MultinomialNB, and GaussianNB, we employed the cross-validation approach.

1. ***Data Biases***

2.1 The Tor Hidden Services.

Users of Tor's hidden service can expose their service without disclosing their address (IP address). Without knowing the service's publisher or disclosing their names, users can connect to it via a rendezvous point. Since the attacker wouldn't be aware of the service's IP address, this kind of anonymity protects against widespread DoS attacks.

Here is the Tor of v11.0. 7 is being used.

2.2 How to reach for search - Anonymous Browsing

For collection of dark web crawling

**5) Quantum Classification Model Discussion**

Google released a pre-trained version of the Universal Sentence Encoder (USE) model on TensorFlowHub, as mentioned in [16] for semantic similarity, text classification, clustering, and other tasks using natural language, the Universal Sentence Encoder converts text into high-dimensional vectors.

The model is developed and tuned for material longer than a word, such as sentences, phrases, or brief paragraphs. With the goal of dynamically accommodating a wide range of natural language comprehension tasks, it is trained on a variety of data sources and workloads. A 512-dimensional vector is produced from English text with variable length as the input.

According to the transfer learning paradigm, this model could be used as a component of a larger network. It could also be used just as-is with its parameters fixed, or it can be fine-tuned by letting the optimized alter them. The output of the USE model is then transferred to subsequent levels to train the network as a whole for a specific downstream job. Document categorization is most likely the most intuitive. In this section, we will transform the abstract of a collection of words retrieved from URLs belonging to distinct categories to sentence embeddings and attach a dense layer with two outputs. Because each output node corresponds to a single category, a softmax activation function is applied before the output is compared to the real labels.

By employing the softmax function, a vector of K real values can be changed into the vector of K real values that equaled 1. The softmax transforms input values—which could be positive, negative, zero, or greater than one—into values between 0 and 1, which can be interpreted as probabilities. Although it will always lie between 0 and 1, the softmax translates small or negative inputs into small probabilities and large or positive inputs into high probabilities.

The Mathematical Definition of Softmax function

|  | The softmax function's input vector is built from of (Z0, ... Zk) |
| --- | --- |
|  | The softmax function's input vector contains all of the Zi values, and they can all have a real value of either a positive, zero or negative sign. The softmax is required because, for instance, a neural network's output vector may be (-032, 4.12, 6.47), which is not really a legitimate probability distribution. |
|  | Every component of an input vector is subjected to the usual exponential function. This results in a positive number that is greater than zero, which will be extremely little if indeed the input was negative and very huge if the input was vast. It is not set within range (0, 1), which is what a probability must be. |
|  | The normalization term is the term in the formula that appears at the bottom. It guarantees that the function's output values will all add up to 1 and fall inside the range (0, 1), forming a legitimate probability distribution. |
| K | how many classes the multi-class classifier can handle. |

After applying the Softmax function with approx 2,000 samples, such a model may be trained in a matter of seconds and achieve 99.6 percent accuracy.

To "quantize" our model, we must substitute a variational quantum circuit for the layer between the embeddings and the output. A traditional dense layer typically has N inputs and M outputs, therefore internally it corresponds to matrix multiplication followed by bias addition and application of the activation function.

However, quantum layers cannot accomplish this openly and must be treated instead. This implies that a quantum variational layer is composed of three processes.

* A traditional dense layer that converts N inputs to N qubits and scales the input by π/2. (so it can represent a rotation around the Bloch sphere)
* A quantum layer that modifies the qubits' states.
* A traditional dense layer that converts the dimensions of N qubits to N output.

To do this with the PennyLane library, one must first define a device that will perform the quantum operations. The real devices are provided by [IBM](https://www.ibm.com/quantum-computing/?p1=Search&p4=p50385922780&p5=e&cm_mmc=Search_Google-_-1S_1S-_-WW_NA-_-ibm%20quantum%20experience_e&cm_mmca7=71700000061253574&cm_mmca8=kwd-397733481108&cm_mmca9=EAIaIQobChMI-bP58eSI5wIVGKSzCh13BACoEAAYASAAEgJE_PD_BwE&cm_mmca10=409646905863&cm_mmca11=e&gclid=EAIaIQobChMI-bP58eSI5wIVGKSzCh13BACoEAAYASAAEgJE_PD_BwE&gclsrc=aw.ds)Q or [Rigetti](https://www.rigetti.com/). Then, the actual circuit is encoded in a python function.

* The PennyLane Library.

A cross-platform open-source software development kit for differentiable quantum computer programming. PennyLane functions by combining machine learning libraries with quantum simulators and hardware, allowing users to train quantum circuits.

Classical calculations, such as model optimization or training, are carried out using typical scientific computing or machine learning libraries, such as SciPy in Python. PennyLane interfaces with these libraries and integrates them with quantum simulators to offer a bridge between classical and quantum computing.

In the same way that a lot of dense layers may be stacked on top of each other to enhance the depth of a network, quantum variational layers can do the same.

* Pseudo Code

Function to create single-layer Hadamard gates taking the number of qubits as parameter:

return performing a loop in range upto n qubits creating Hadamard layer

Function to create a layer of parameterized qubit rotation around y-axis taking feature as parameter:

return the rotated qubits up to certain angle provided in the parameter

Function to entangle the layers by taking the number of qubits as parameter:

For i to nth qubit and traverse in 2 steps each:

Get the even index to add CNOT gate

For j to nth qubit and iterate with 2 steps each:

Get the odd index to add CNOT gate

Class for VariationalQuantumCircuit:   
 constructor taking input as n\_categories to classify the URL,

Number of qubits required (n\_qubits = 4),

Layers of circuit need = 6

Function to create a circuit taking parameters as inputs and parameters:

Embedding encode to Setting the templates from features into the quantum state of the circuit.

Setting up the layer for StronglyEntanglingLayers due to which takes parameter as number of list in a dataset.

return expected value of Pauli Z for i in range of n\_qubits

Function to create the quantm Circuit:

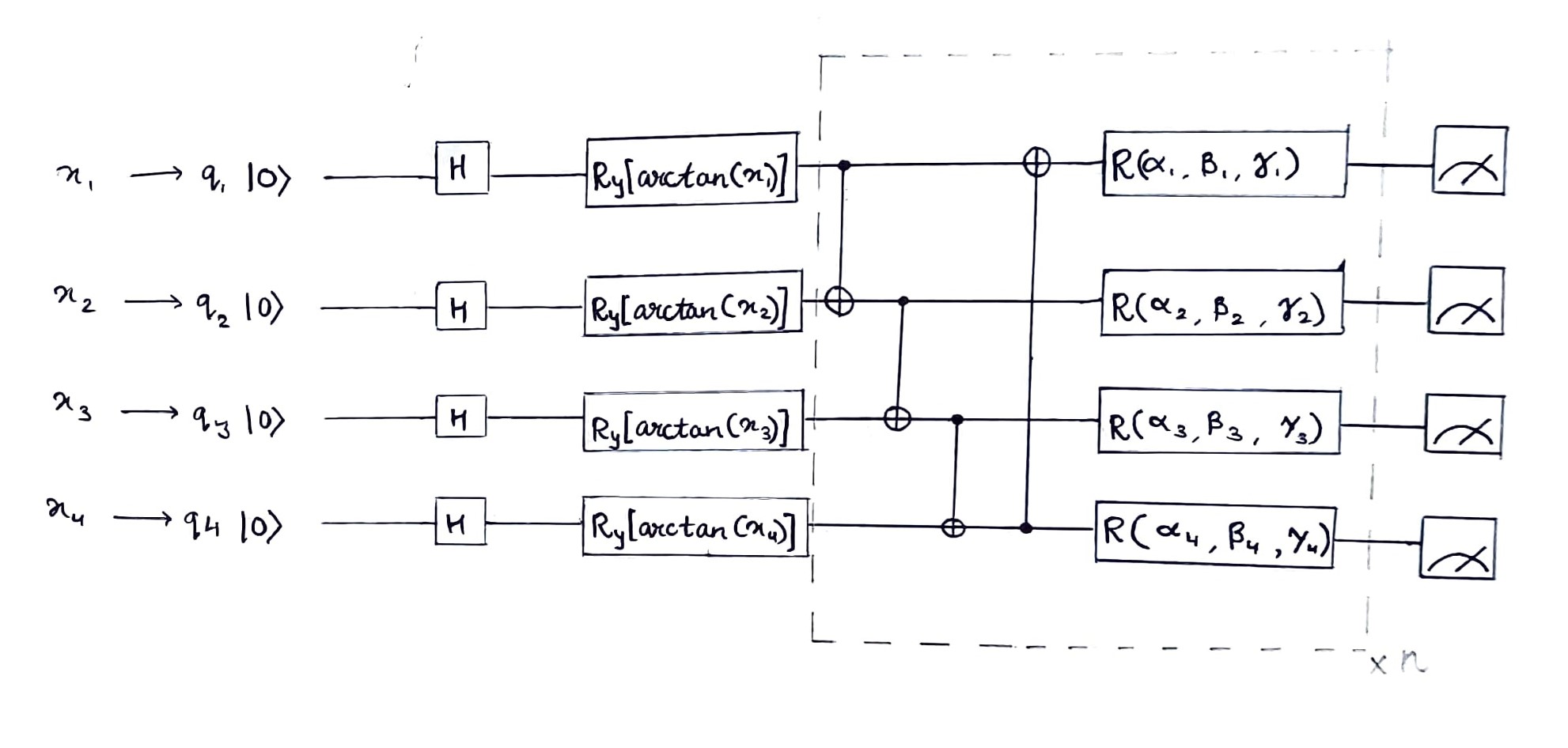
UniversalEmbeddingLayer= partial(USELsayer)

Initiating the quantum circuit variational circuit object

x = softmax function activation

Return denselayer.classification

* *Quantum Classification Variational Circuit*

**

* *An impartial beginning state is created using the Hadamard gate H.*
* *The three sections of the VQC (Variational Quantum Circuit) component employed in this study are the encoding portion, the variational part containing parameters to be learned, and the measurement part, which will provide the Pauli-Z expectation values through repeated quantum circuit runs. The first k qubits would be used for the quantum measurement.*
* *For state preparation, arctan(x) is used, and they are taken straight from the input classical data (cleaned words scrapped from HTML).*
* *The number of classes is k. The dashed section is iterated n times to boost the VQC's expressivity.*
* *x1, x2, x3, x4 represents the encoding of classical data into Quantum Data by Universal Sentence Encoder,*
* *Before starting the circuit all the Qubits are brought to their Zeroth State.*
* *The Number of Qubits chosen depends on the input dimension of the data and the depth of circuit we need (eg here 512 dimensions)*
* **Result and Discussion**

The quantum version of the classifier takes substantially longer to train using the simulator (approximately one minute per epoch with a batch size of 32); if we had efficient simulations of quantum processes, we would not require quantum hardware in the first place! In reality, no such entity can exist in theory since quantum physics possesses qualities like entanglement that do not exist in the classical universe. The major point is that simulations of quantum processes are time-consuming and grow progressively unmanageable as the number of qubits increases.

* **Limitation and Future Scope**

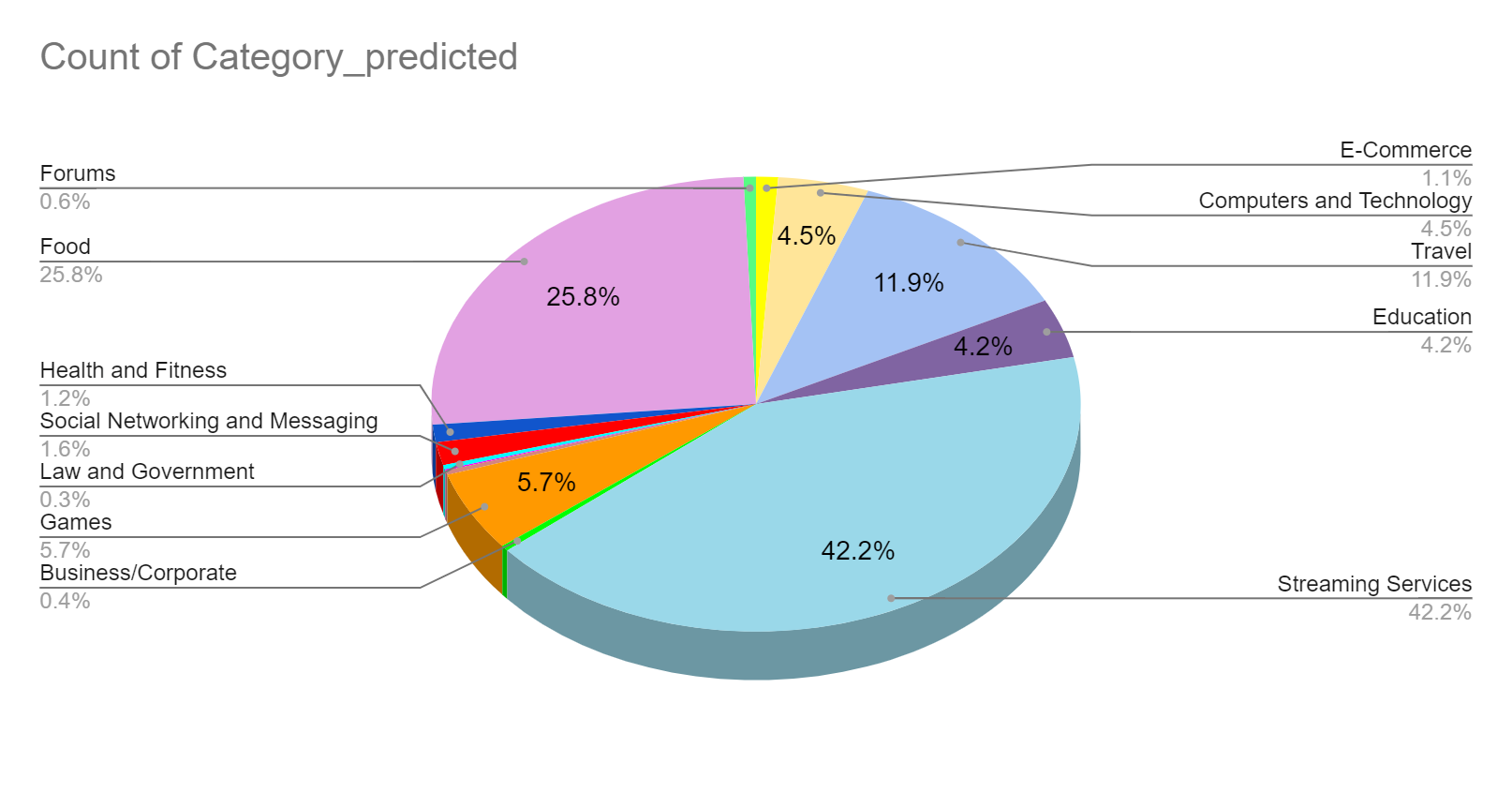
As mentioned in [13], Real-world quantum hardware, on the other hand, is noisy, as defined by John Preskill as the Noisy Intermediate-Scale Quantum (NISQ) technological period. In reality, thermal noise inhibits the capacity to do extended calculations without introducing mistakes and is likely the primary impediment to producing large-scale, trustworthy quantum technology today.

However, compared to the purely classical model, the hybrid network can execute classification with an accuracy of 98.8%. This should also not come as a huge surprise as the USE embedding layer is actually the one doing the heavy labor. This is a noteworthy conclusion, though, as it allows us to employ transfer learning rather than having to train a network that is only capable of doing quantum computations for the same job.

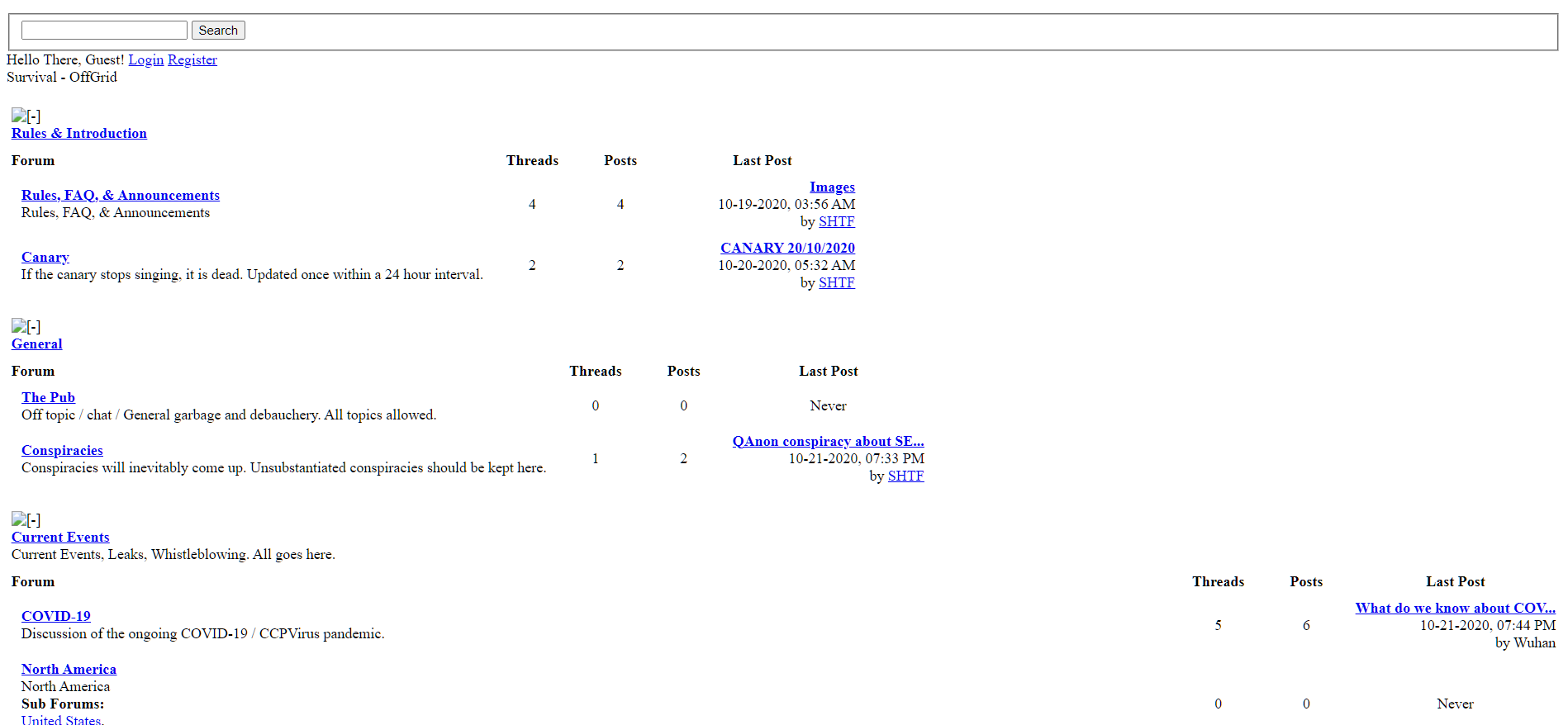
* **Sample Output**

| **Id** | **URLs** | **Title** | **Category\_Predicted** |
| --- | --- | --- | --- |
| 1 | 4p6i33oqj6wgvzgzczyqlueav3tz456rdu632xzyxbnhq4gpsriirtqd.onion/3CB6NF7FCFC02A2CFB2C8E7EFBAF8E9.html | Peoples Drug Store-The Darkweb's Best Online Drug Supplier! - Buy cocaine, speed, xtc, mdma, heroin and more at peoples drug store, pay with Bitcoin | E-Commerce |
| 2 | canxzwmfihdnn7bz.onion/A7045E8B1F9E04E6GF35DF8b96C29.html | Ableonion | Social Network & Messaging |
| 3 | crawled\bepig5bcjdhtlwpgeh3w42hffftcqmg7b77vzu7ponty52kiey5ec4ad.onion\F7E89E9EDFFByAKxE8c3E99C6s.html | Kamagra For Bitcoin - Same quality as original viagra pills, cheap prices, Bitcoin payment | Health & Fitness |
| 4 | crawled\ar-ar.facebookcorewwwi.onion\6ADH89JBD10FB2A1B1084DABFF424.html | ‪Belarus Solidarity Foundation‬ | Law & Government |
| 5 | crawled\ctemplarpizuduxk3fkwrieizstx33kg5chlvrh37nz73pv5smsvl6ad.onion\0F2AE691K5B841915B81BED0743CEA.html | The Only Anonymous Payment Resources You Will Ever Need? - CTemplar | Business/Corporate |
| 6 | crawled\cx-ph.facebookcorewwwi.onion\BDF9CA93A50CD09FY9C2AkF441DG.html | Announcements | Facebook Help Center | Facebook | Streaming Services |
| 7 | crawled\zqktlwiuavvvqqt4ybvgvi7tyo4hjl5xgfuvpdf6otjiycgwqbym2qad.onion\YD68EB1A0aB34823CE1700JW.html | Talk:Link proposals - The Hidden Wiki | Education |
| 8 | crawled\bg-bg.facebookcorewwwi.onion\0BDB8C9ECFDF1A8DA7-11FE007Bl.html | Facebook Fundraising Campaigns Category... | Forums |
| 9 | crawled\ca-es.facebookcorewwwi.onion\A029BF3B3287BCEzAD29989F6F086.html | Aspectos básicos de la privacidad de Facebook | Food |
| 10 | crawled\da-dk.facebookcorewwwi.onion\ACFU9C\_-0AC61298614DD2A77B.html | Local Guides | Travel |
| 11 | crawled\de-de.facebookcorewwwi.onion\0Aa03F61DBFE6i5D1960F867BCE.html | Veg-Friendly Delights | Games |
| 12 | crawled\zqktlwiuavvvqqt4ybvgvi7tyo4hjl5xgfuvpdf6otjiycgwqbym2qad.onion\YB1D849901dB6F9d8AEE048BD1l.html | Network Attributes of your computer - The Hidden Wiki | Computer & Technology |

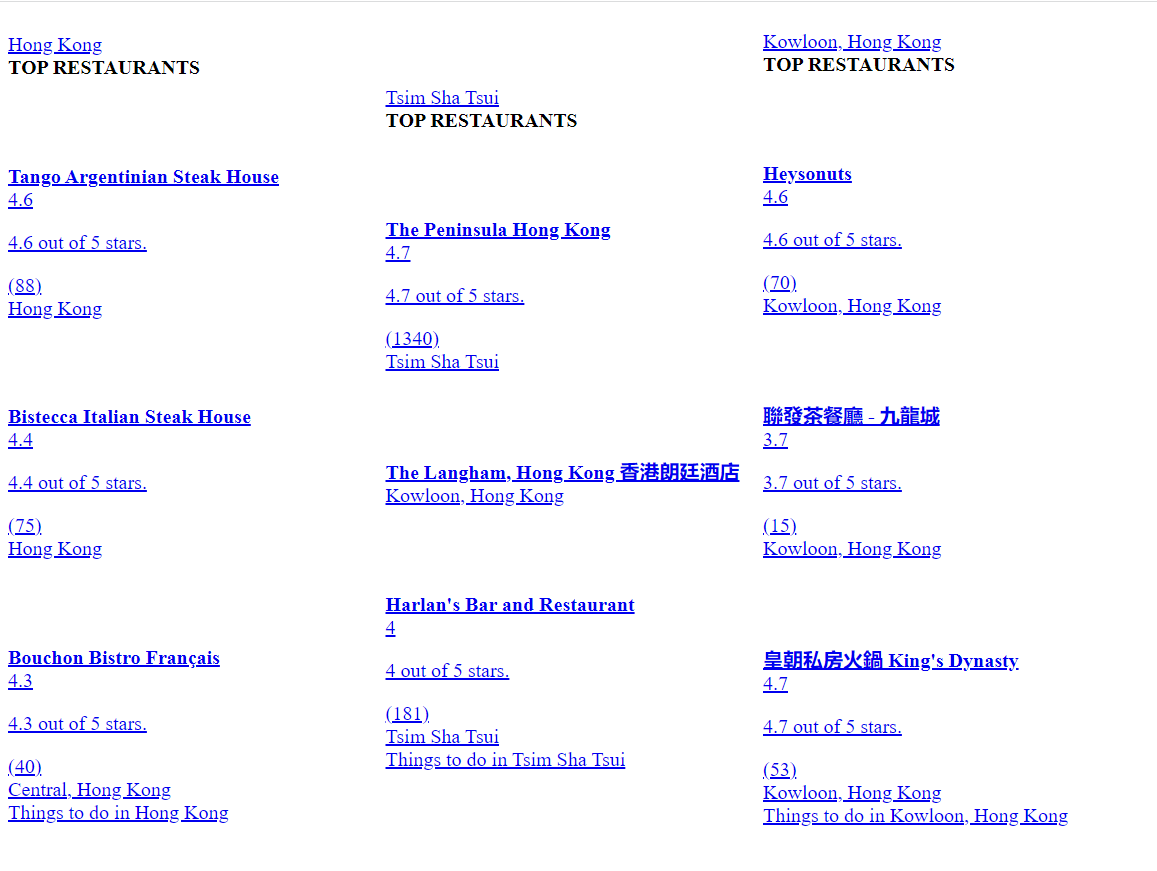
**For 2000 URLs the classification is represented as:**

****

**Category**- Forums



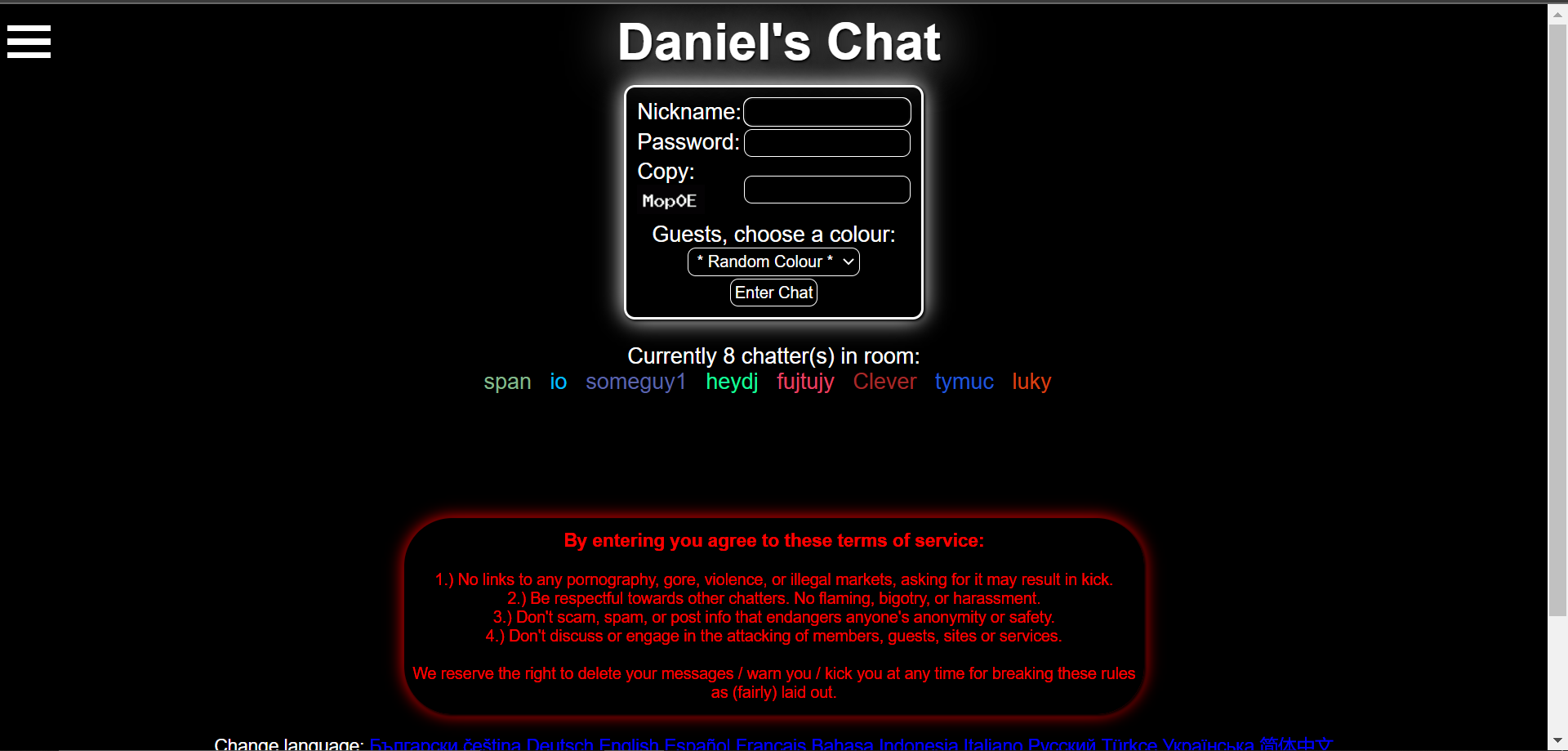
**Category**- Food



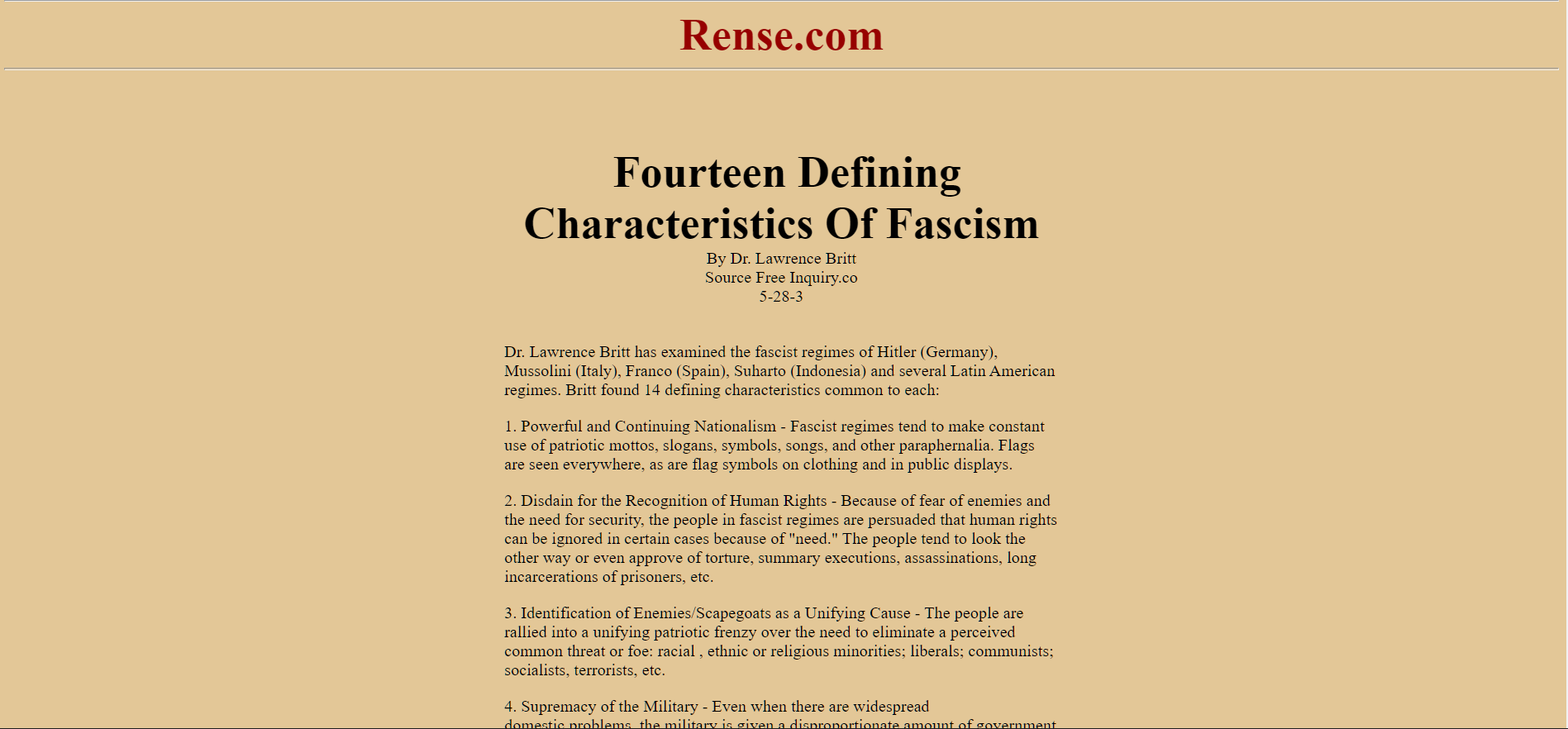
**Category**- Health & Fitness



**Category**- Social Networking & Messaging



**Category**- Law & Government



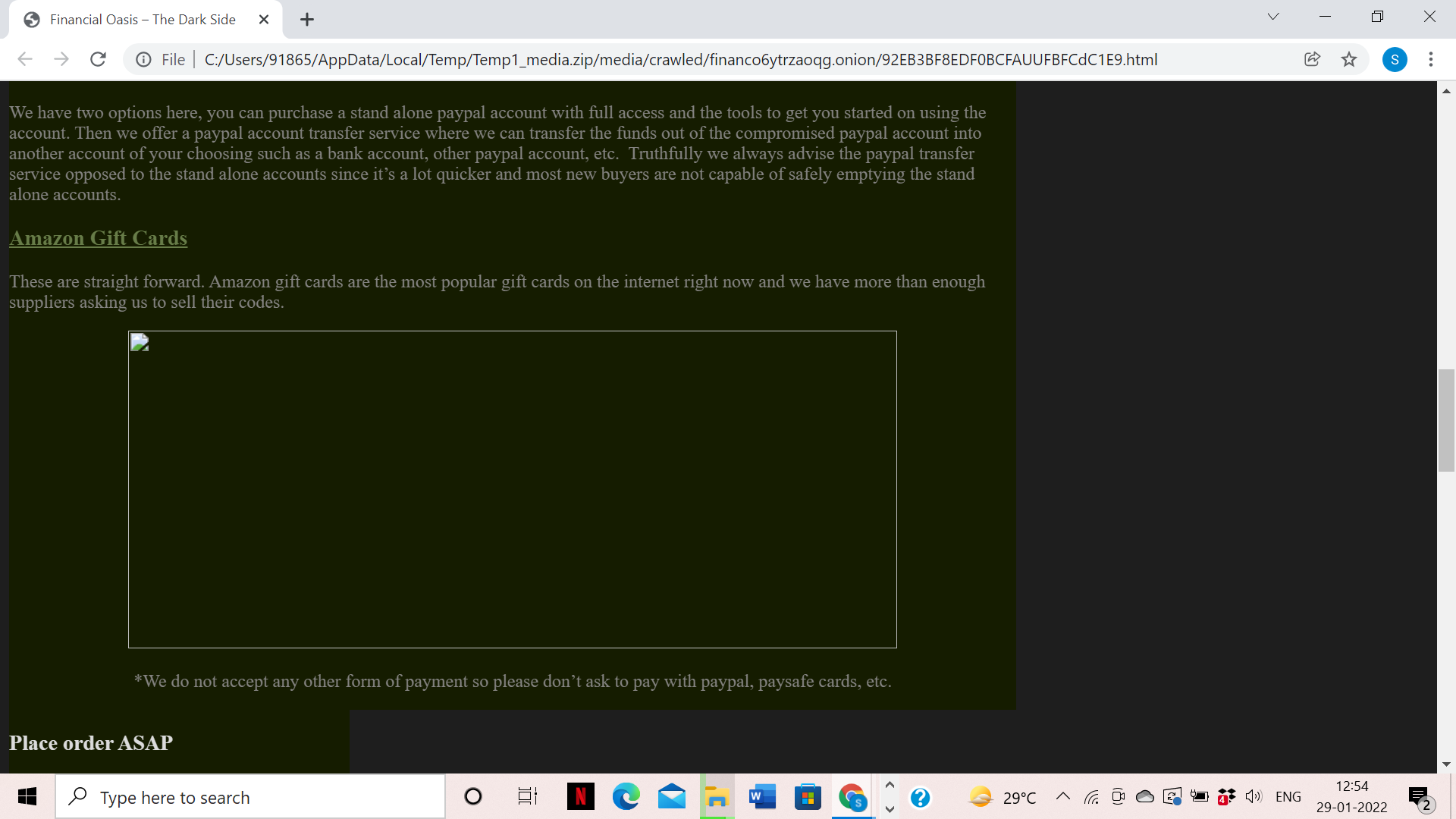
**Category**- Games



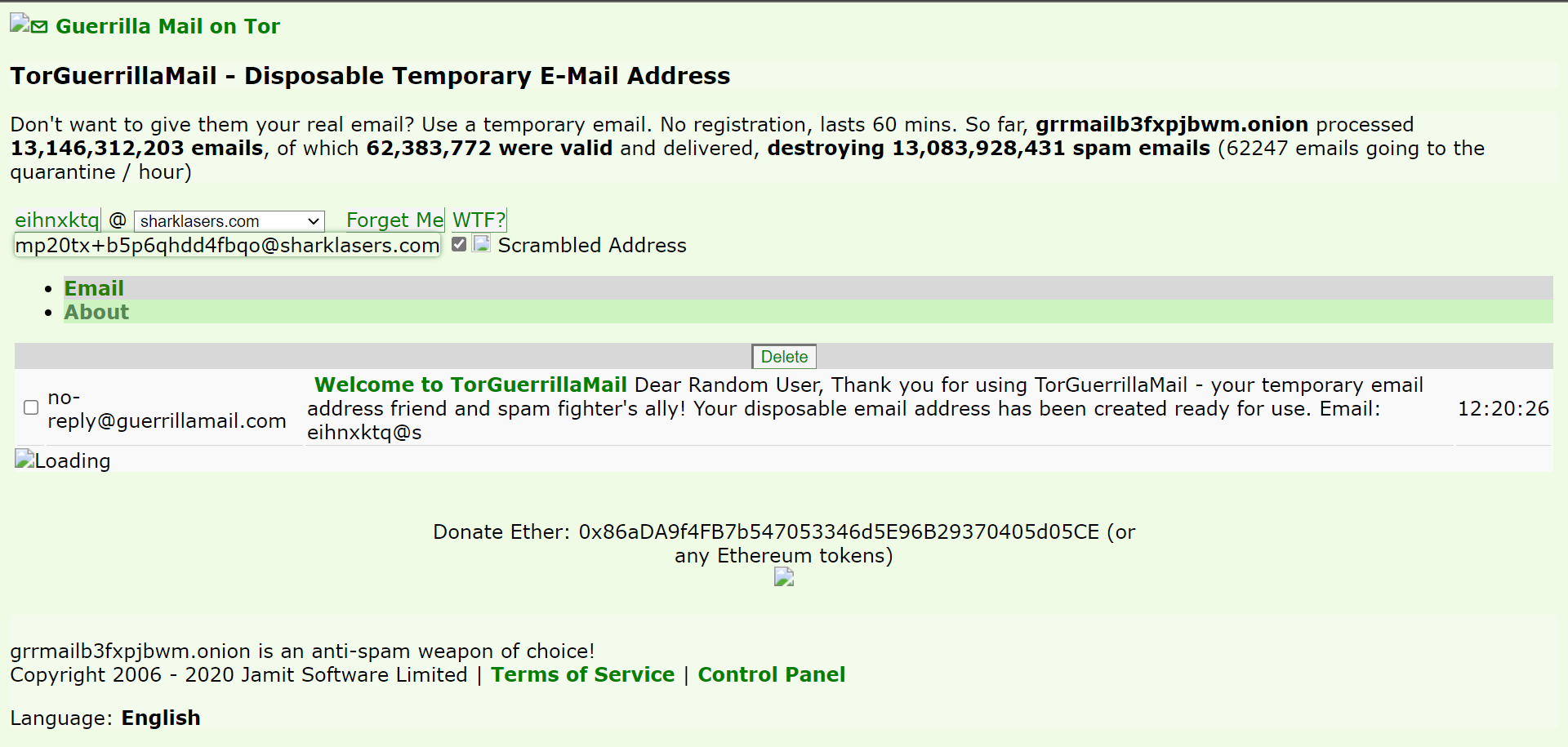
**Category**- Business/Corporate



**Category**- E-Commerce



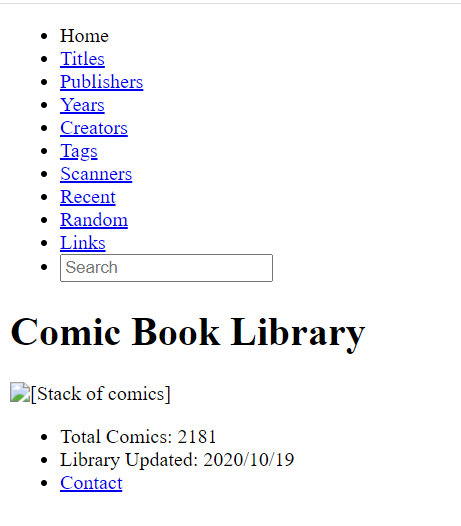
**Category**- Computers and Technology



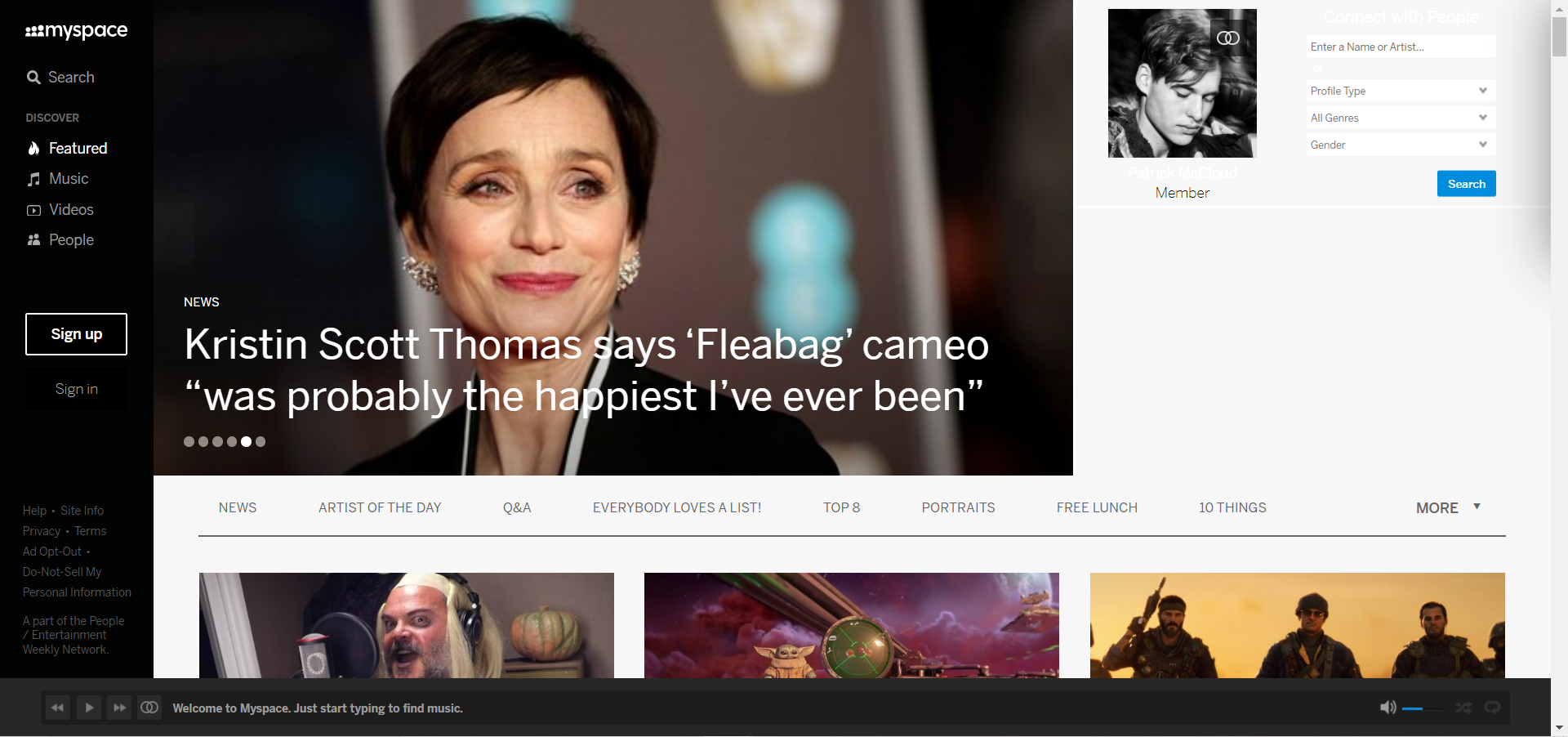
**Category**- Travel



**Category**- Education



**Category**- Streaming Services



* **Performance Graph of Classical Model and Quantum Model**

| **Fig1. Classical Model: Time Required Vs Batch size of 2000 URLs**  **Fig2. Quantum Model: Time Required Vs Batch Size of 2000 URLs** |
| --- |

**In the Time performance graph:**

**Classical Model:** the incoming cleaned words have different sizes, for example as URL A has 5000 words where URL B has 7500 words, here the classical model takes comparatively more time to categorize the URL B. As the batch size increases which in turn increases the overall categorization time.

**Quantum Model:** the incoming cleaned words from the classical model are converted to quantum data which has less memory as qubits can represent an exponential number of bits, the overall time slightly increases when we have more than 8000 words that are not able to represent the whole data in qubits. The words which don't fit in a single interaction (processing of a single URL) are discarded. Also in the end before the final output, the data is passed through the Softmax function so as to ensure the output value is correct/matched.

| **Fig3. Classical Model: Memory Required Vs Batch size of 2000 URLs**    **Fig4. Quantum Model: Time Required Vs Batch Size of 2000 URLs** |
| --- |

**In the Memory performance graph:**

**Classical Model:** the incoming cleaned words have different sizes, for example as URL A has 5000 words where URL B has 7500 words, here the memory keeps on increasing thus slowing down the time taken to classify the URL.

**Quantum Model:** the incoming cleaned words from the classical model are converted to quantum data which has less memory as qubits can represent an exponential number of bits. In every Iteration (classifying the 1 URL at a time) the qubits are conserved (before and after the output the number of qubits involved in the circuit is equal) and the value of words that are less significant in classification stays in the qubit for a long time until the next words are encoded into the same qubit.

* **Conclusion**

This research work to Improve the classical model of URL Classification by Classifying the extracted keywords from URL classifying with Quantum enhanced Transfer Learning.

We extracted the data from URLs and successfully fetched the data and encoded it into Quantum data via Universal Sentence Encoder and Softmax Function and implemented the model on Quantum Circuit, Hence we get the Category of the URL.

* **References**

1. Cer, D., Yang, Y., Kong, S.Y., Hua, N., Limtiaco, N., John, R.S., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C. and Sung, Y.H., 2018. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*.
2. Abohashima, Z., Elhosen, M., Houssein, E.H. and Mohamed, W.M., 2020. Classification with quantum machine learning: A survey. *arXiv preprint arXiv:2006.12270*.
3. Schuld, M. and Petruccione, F., 2018. Quantum ensembles of quantum classifiers. *Scientific reports*, *8*(1), pp.1-12.
4. Mitarai, K., Negoro, M., Kitagawa, M. and Fujii, K., 2018. Quantum circuit learning. *Physical Review A*, *98*(3), p.032309.
5. Ramezani, S.B., Sommers, A., Manchukonda, H.K., Rahimi, S. and Amirlatifi, A., 2020, July. Machine learning algorithms in quantum computing: A survey. In *2020 International joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.
6. Zheng, Y., Lu, S. and Wu, R.B., 2018. Quantum circuit design for training perceptron models. *arXiv preprint arXiv:1802.05428*.
7. Sim, S., Johnson, P.D. and Aspuru‐Guzik, A., 2019. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum‐classical algorithms. *Advanced Quantum Technologies*, *2*(12), p.1900070.
8. Chalumuri, A., Kune, R. and Manoj, B.S., 2021. A hybrid classical-quantum approach for multi-class classification. *Quantum Information Processing*, *20*(3), pp.1-19.
9. Fastovets, D.V., Bogdanov, Y.I., Bantysh, B.I. and Lukichev, V.F., 2019, March. Machine learning methods in quantum computing theory. In *International Conference on Micro-and Nano-Electronics 2018* (Vol. 11022, pp. 752-761). SPIE.
10. Khoshaman, A., Vinci, W., Denis, B., Andriyash, E., Sadeghi, H. and Amin, M.H., 2018. Quantum variational autoencoder. *Quantum Science and Technology*, *4*(1), p.014001.
11. Sharma, S., 2020. Qeml (quantum-enhanced machine learning): Using quantum computing to enhance ml classifiers and feature spaces. *arXiv preprint arXiv:2002.10453*.
12. Cortese, J.A. and Braje, T.M., 2018. Loading classical data into a quantum computer. *arXiv preprint arXiv:1803.01958*.
13. Preskill, J., 2018. Quantum computing in the NISQ era and beyond. *Quantum*, *2*, p.79.
14. Mosca, M., 2018. Cybersecurity in an era with quantum computers: will we be ready?. *IEEE Security & Privacy*, *16*(5), pp.38-41.
15. Ko, K.K. and Jung, E.S., 2021. Development of cybersecurity technology and algorithm based on quantum computing. *Applied Sciences*, *11*(19), p.9085.
16. Kilber, N., Kaestle, D. and Wagner, S., 2021. Cybersecurity for Quantum Computing. *arXiv preprint arXiv:2110.14701*.
17. Espitia, G., 2021. On the Role of Quantum Computing in Science and Cybersecurity. *arXiv preprint arXiv:2105.09942*.
18. Vaishnavi, A. and Pillai, S., 2021, July. Cybersecurity in the Quantum Era-A Study of Perceived Risks in Conventional Cryptography and Discussion on Post Quantum Methods. In *Journal of Physics: Conference Series* (Vol. 1964, No. 4, p. 042002). IOP Publishing.
19. Fioravanti, T., 2022. Evaluation of quantum machine learning algorithms for cybersecurity.
20. Ramezani, S.B., Sommers, A., Manchukonda, H.K., Rahimi, S. and Amirlatifi, A., 2020, July. Machine learning algorithms in quantum computing: A survey. In *2020 International joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.
21. Gianani, I., Mastroserio, I., Buffoni, L., Bruno, N., Donati, L., Cimini, V., Barbieri, M., Cataliotti, F.S. and Caruso, F., 2021. Experimental quantum embedding for machine learning. *Advanced Quantum Technologies*, p.2100140.
22. Garg, D., Ikbal, S., Srivastava, S.K., Vishwakarma, H., Karanam, H. and Subramaniam, L.V., 2019. Quantum embedding of knowledge for reasoning. *Advances in Neural Information Processing Systems*, *32*.
23. Kübler, J.M., Muandet, K. and Schölkopf, B., 2019. Quantum means embedding of probability distributions. *Physical Review Research*, *1*(3), p.033159.
24. Broughton, M., Verdon, G., McCourt, T., Martinez, A.J., Yoo, J.H., Isakov, S.V., Massey, P., Halavati, R., Niu, M.Y., Zlokapa, A. and Peters, E., 2020. Tensorflow quantum: A software framework for quantum machine learning. *arXiv preprint arXiv:2003.02989*.
25. Nath, R.K., Thapliyal, H. and Humble, T.S., 2021. A review of machine learning classification using quantum annealing for real-world applications. *SN Computer Science*, *2*(5), pp.1-11.
26. Peters, E., Caldeira, J., Ho, A., Leichenauer, S., Mohseni, M., Neven, H., Spentzouris, P., Strain, D. and Perdue, G.N., 2021. Machine learning of high dimensional data on a noisy quantum processor. *npj Quantum Information*, *7*(1), pp.1-5.
27. Trukhanov, A., Platonov, A. and Bessmertny, I., 2020. Using quantum probability for word embedding problem. In *CEUR Workshop Proceedings* (Vol. 2590).
28. Liang, J.M., Shen, S.Q., Li, M. and Li, L., 2020. Variational quantum algorithms for dimensionality reduction and classification. *Physical Review A*, *101*(3), p.032323.
29. He, S., He, Y. and Li, M., 2019, March. Classification of illegal activities on the dark web. In *Proceedings of the 2019 2nd International Conference on Information Science and Systems* (pp. 73-78).
30. Alaidi, A.H.M., Al\_airaji, R.A.M., ALRikabi, H.T., Aljazaery, I.A. and Abbood, S.H., 2022. Dark Web Illegal Activities Crawling and Classifying Using Data Mining Techniques. *International Journal of Interactive Mobile Technologies*, *16*(10).
31. Park, J., Mun, H. and Lee, Y., 2018, December. Improving tor hidden service crawler performance. In *2018 IEEE Conference on Dependable and Secure Computing (DSC)* (pp. 1-8). IEEE.
32. Zabihimayvan, M. and Doran, D., 2022. A first look at references from the dark to the surface web world: a case study in Tor. *International Journal of Information Security*, pp.1-17.
33. Nair, V. and Kannimoola, J.M., 2022. A Tool to Extract Onion Links from Tor Hidden Services and Identify Illegal Activities. In *Inventive Computation and Information Technologies* (pp. 29-37). Springer, Singapore.
34. Buldin, I.D. and Ivanov, N.S., 2020, January. Text classification of illegal activities on onion sites. In *2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)* (pp. 245-247). IEEE.
35. Avarikioti, G., Brunner, R., Kiayias, A., Wattenhofer, R. and Zindros, D., 2018. Structure and content of the visible Darknet. *arXiv preprint arXiv:1811.01348*.
36. He, S., He, Y. and Li, M., 2019, March. Classification of illegal activities on the dark web. In *Proceedings of the 2019 2nd International Conference on Information Science and Systems* (pp. 73-78).