IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

An Efficient Deep Learning Technique for Detecting of Spam in IOT network

Sarfaraj Alam[1], Ms. Sonal Chaudhary[2]

M.Tech Scholar[1], Associate Professor[2]

Department of Computer Science & Engineering

All Saints' College of Technology, Bhopal, Madhya Pradesh, India

Abstract— **The massive number of sensors deployed in the Internet of Things (IoT) produce gigantic amounts of data for facilitating a wide range of applications. Deep Learning (DL) would undoubtedly play a role in generating valuable inferences from this massive volume of data and hence will assist in creating smarter IoT. Spamming is the use of messaging or electronic messaging system that send huge amount of data. Spam often fills the internet with multiple copies of a message and are sent to different recipients repeatedly without their request and urges to open them. Spam is type of virus, it is sent for commercial purposes. It can be sent in massive volume by botnets, networks of infected computers. This paper proposed efficient deep learning technique of spam detection for IOT devices application. The simulation is performed using the Python Spyder Software.**

Keywords—Spam. IOT, Python, Deep Learning, Detection, Virus.

I. INTRODUCTION

Internet of Things (IoT) enables convergence and implementations between the real-world objects irrespective of their geographical locations. Implementation of IOT network management and control makes privacy and protection strategies utmost important and challenging in such an environment. IoT applications need to protect data privacy to fix security issues such as intrusions, spoofing attacks, DoS attacks, DoS attacks, jamming, eavesdropping, spam, and malware.

The safety measures of IoT devices depend upon the size and type of organization in which it is imposed. The behavior of users forces the security gateways to cooperate. In other words, we can say that the location, nature, application of IoT devices decides the security measures. For instance, the smart IoT security cameras in the smart organization can capture the different parameters for analysis and intelligent decision making. The maximum care to be taken is with web based devices as maximum number of IoT devices are web dependent. It is common at the workplace that the IoT devices installed in an organization can be used to implement security and privacy features efficiently. For example, wearable devices collect and send user's health data to a connected smartphone should prevent leakage of information to ensure privacy. It has been found in the market that 25-30% of working employees connect their personal IoT devices with the organizational network. The expanding nature of IoT attracts both the audience, i.e., the users and the attackers.

However, with the emergence of ML in various attacks scenarios, IoT devices choose a defensive strategy and decide the key parameters in the security protocols for trade-off between security, privacy and computation. This job is challenging as it is usually difficult for an IoT system with limited resources to estimate the current network and timely attack status.

Denial of service (DDoS) attacks: The attackers can flood the target database with unwanted requests to stop IoT devices from having access to various services. These malicious requests produced by a network of IoT devices are commonly known as bots. DDoS can exhaust all the resources provided by the service provider. It can block authentic users and can make the network resource unavailable. These are the attacks imposed at the physical layer of IoT device. This attack leads to lose the integrity of the device. Attackers attempt to modify the data either at the node storage or while it is in the transmission within network. The common attacks possible at the sensor node are attacks on availability, attacks on authenticity, attacks on confidentiality, Cryptography keys brute-forcing. The countermeasures to ensure prevention of such attacks includes password protection, data encryption and restricted access control.

The Internet of Things (IoT) is a group of millions of devices having sensors and actuators linked over wired or wireless channel for data transmission. IoT has grown rapidly over the past decade with more than 25 billion devices expected to be connected by 2020. The volume of data released from these devices will increase many-fold in the years to come. In addition to an increased volume, the IoT devices produces a large amount of data with a number of different modalities having varying data quality defined by its speed in terms of time and position dependency. In such an environment, machine learning (ML) algorithms can play an important role in ensuring security and authorization based on biotechnology, anomalous detection to improve the usability, and security of IoT systems [1].

II. LITERATURE SURVEY

A. Makkar et al.,[1] this framework, five ML models are evaluated using various metrics with a large collection of inputs features sets. Each model computes a spam score by considering the refined input features. This score depicts the trustworthiness of IoT device under various parameters. REFIT Smart Home data set is used for the validation of proposed technique. The results obtained prove the effectiveness of the proposed scheme in comparison to the other existing schemes.

F. Hossain et al.,[2] proposed a model that classifies the e-mail into spam and ham. DBSCAN and Isolation Forest are used to identify the extreme values outside of the specific range. Heatmap, Recursive Feature Elimination, and Chi-Square feature selection techniques are used to select the effective features. The proposed model is implemented in both machine learning and deep learning to establish a comparative analysis.

A. Makkar et al.,[3] spammer framework for web spam detection is proposed. CSF detects the web spam by fuzzy rule based classifiers along with machine learning classifiers. Each classifier produces the quality score of the webpage. These quality scores are then ensembled to generate a single score, which predicts the spamicity of the web page. For ensembling, fuzzy voting approach is used in CSF.

G. Fortino et al.,[4] use of a reputation model can be a practicable and effective solution to form local communities of agents on the basis of their social capabilities. In this work, we propose a framework for agents operating in an IoT environment, called ResIoT, where the formation of communities for collaborative purposes is performed on the basis of agent reputation.

K. A. Al-Thelaya et al.,[5] graph-based datasets. The representation models are mainly developed based on analyzing interactions and relations between users. The first model is developed based on graph-based analysis, while the other one is developed based on sequential processing of user interactions. Based on the conducted experiments, we conclude that the two representation models show high spam detection accuracy.

T. Y. Ho et al.,[6] In recent years, insider threats within computers have been overgrowing because a high quantity of malware and its variants have been spread massively by spam mail, malvertising attack, and users' carelessness. Moreover, some of the dormant malware would not be inspected by ant-virus software, and the risk exists continually until finally becoming a disastrous economic loss. Several studies developed signature-based methods to detect insider threats, but we are more interested in how to simplify the network behavior from sophisticated traffic flow.

J. Zhang et al.,[7] proposes a method based on Wasserstein Generative Adversarial Network(WGAN) to generate malicious PDF files which are similar to benign ones and can evade the malicious file detection system. The experimental results show that the adversarial examples generated by our method can evade the PDF classifier-PDFrate of 100%.

A. Makkar et al.,[8] spam webpages are detected before these are processed by the ranking module of search engines. The machine learning model, i.e., decision tree is used for the validation of the proposed scheme. The ten fold cross validation approach is used to improve the accuracy of model, i.e., 98.2%. The results obtained demonstrate that the proposed scheme has the power of preventing the spam web pages in Cognitive Internet of Things (CIoT) environment.

A. K. Singh et al.,[9] analyze different machine learning techniques with feature selection and without feature selection algorithms and their performance to detect the best classifier for spam mail classification. First, we apply each classifier without selecting any features in order to experiment on the dataset and examine the outcome. Next, to select the desired features we apply best first feature selection algorithm and apply various algorithms for classification.

T. Lange et al.,[10] network topology and technology undergirding each botnet varies greatly, as do the motivations commonly behind such networks. Furthermore, as botnets have continued to evolve, many newer ones demonstrate increased levels of anonymity and sophistication, making it more difficult to effectively counter them. Increases in the production of vulnerable Internet of Things (IoT) devices has made it easier for malicious actors to quickly assemble sizable botnets.

T. Qiu et al.,[11] purpose of benefit maximization, which has caused confusion and heavy losses to industrial production. It is difficult to distinguish spammers from normal users owing to the characteristics of multidimensional data. To address this problem, this work proposes a spammer identification scheme based on Gaussian mixture model (SIGMM) that utilizes machine learning for industrial mobile networks.

G. Kumar et al.,[12] Social Networking plays a very important part in our day to day life to share our views on various issues those arise. As it opened new means of communication between masses to share their thoughts. The data from these sites can be very useful for the purpose of analysis.

III. METHODOLOGY

The main contributions of this work is to collect SPAM dataset from kaggle website and implement deep learning technique to detection of the spam for the IOT devices application.

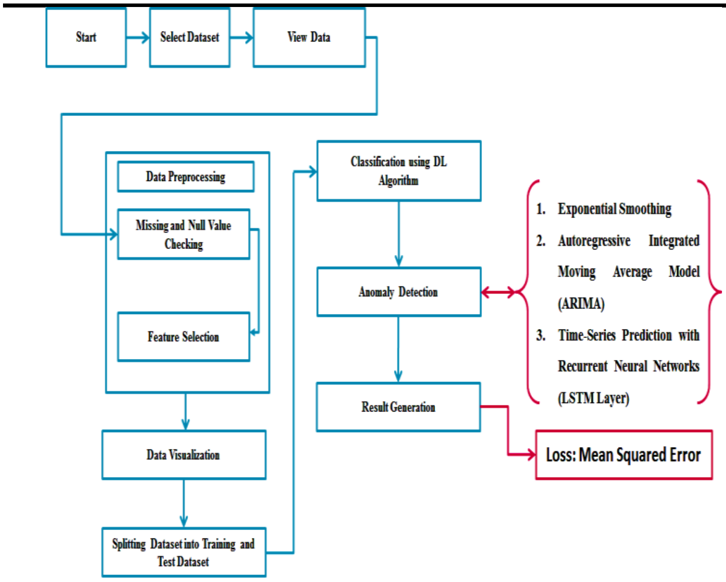

Start
Select Dataset
View Data
Data Preprocessing
Missing and Null Value Checking
Feature Selection
Data Visualization
Splitting Dataset into Training and Test Dataset
Classification using DL Algorithm
Anomaly Detection
Result Generation
1. Exponential Smoothing
2. Autoregressive Integrated Moving Average Model (ARIMA)
3. Time-Series Prediction with Recurrent Neural Networks (LSTM Layer)
Loss: Mean Squared Error

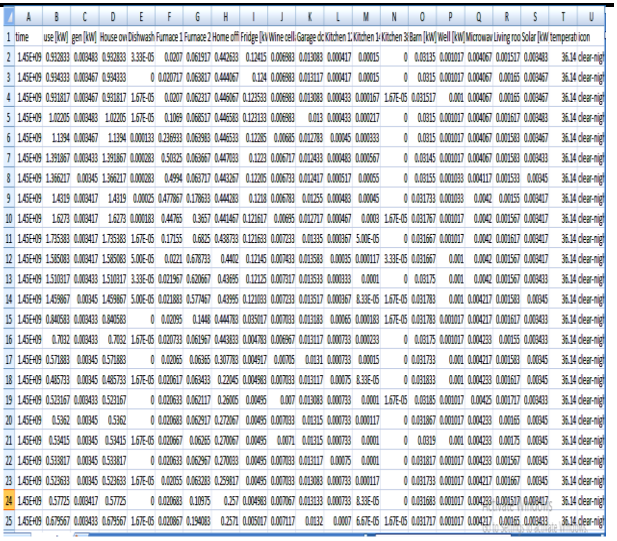
Figure 1: Flow Chart

Steps-

1. Firstly, download the EEG dataset from kaggle website, which is a large dataset provider and machine learning repository Provider Company for research.
2. Now apply the preprocessing of the data, here handing the missing data, removal null values.
3. Now extract the data features and evaluate in dependent and independent variable.
4. Now apply the classification method based on the deep learning (LSTM) approach.
5. Now generate confusion matrix and show all predicted class like true positive, false positive, true negative and false negative.
6. Now calculate the performance parameters by using the standard formulas in terms of the precision, recall, F_measure, accuracy and error rate.

IV. SIMULATION RESULTS

The execution of the proposed calculation is done over python spyder 3.7. The sklearn, numpy, pandas, matplotlib, pyplot, seaborn, os library assists us with utilizing the capacities accessible in spyder climate for different strategies.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	time	use [kW]	gen [kW]	House ov	Dishwash	Furnace 1	Furnace 2	Home offi	Fridge [kV	Wine cell	Garage dc	Kitchen 1	Kitchen 1	Kitchen 3	Barn [kW]	Well [kW]	Microwav	Living roo	Solar [kW	temperat	icon
2	1.45E+09	0.932833	0.003483	0.932833	3.33E-05	0.0207	0.061917	0.442633	0.12415	0.006983	0.013083	0.000417	0.00015	0	0.03135	0.001017	0.004067	0.001517	0.003483	36.14	clear-nig
3	1.45E+09	0.934333	0.003467	0.934333	0	0.020717	0.063817	0.444067	0.124	0.006983	0.013117	0.000417	0.00015	0	0.0315	0.001017	0.004067	0.00165	0.003467	36.14	clear-nig
4	1.45E+09	0.931817	0.003467	0.931817	1.67E-05	0.0207	0.062317	0.446067	0.123533	0.006983	0.013083	0.000433	0.000167	1.67E-05	0.031517	0.001	0.004067	0.00165	0.003467	36.14	clear-nig
5	1.45E+09	1.02205	0.003483	1.02205	1.67E-05	0.1069	0.068517	0.446583	0.123133	0.006983	0.013	0.000433	0.000217	0	0.0315	0.001017	0.004067	0.001617	0.003483	36.14	clear-nig
6	1.45E+09	1.1394	0.003467	1.1394	0.000133	0.236933	0.063983	0.446533	0.12285	0.00685	0.012783	0.00045	0.000333	0	0.0315	0.001017	0.004067	0.001583	0.003467	36.14	clear-nig
7	1.45E+09	1.391867	0.003433	1.391867	0.000283	0.50325	0.063667	0.447033	0.1223	0.006717	0.012433	0.000483	0.000567	0	0.03145	0.001017	0.004067	0.001583	0.003433	36.14	clear-nig
8	1.45E+09	1.366217	0.00345	1.366217	0.000283	0.4994	0.063717	0.443267	0.12205	0.006733	0.012417	0.000517	0.00055	0	0.03155	0.001033	0.004117	0.001533	0.00345	36.14	clear-nig
9	1.45E+09	1.4319	0.003417	1.4319	0.00025	0.477867	0.178633	0.444283	0.1218	0.006783	0.01255	0.000483	0.00045	0	0.031733	0.001033	0.0042	0.00155	0.003417	36.14	clear-nig
10	1.45E+09	1.6273	0.003417	1.6273	0.000183	0.44765	0.3657	0.441467	0.121617	0.00695	0.012717	0.000467	0.0003	1.67E-05	0.031767	0.001017	0.0042	0.001567	0.003417	36.14	clear-nig
11	1.45E+09	1.735383	0.003417	1.735383	1.67E-05	0.17155	0.6825	0.438733	0.121633	0.007233	0.01335	0.000367	5.00E-05	0	0.031667	0.001017	0.0042	0.001617	0.003417	36.14	clear-nig
12	1.45E+09	1.585083	0.003417	1.585083	5.00E-05	0.0221	0.678733	0.4402	0.12145	0.007433	0.013583	0.00035	0.000117	3.33E-05	0.031667	0.001	0.0042	0.001567	0.003417	36.14	clear-nig
13	1.45E+09	1.510317	0.003433	1.510317	3.33E-05	0.021967	0.620667	0.43695	0.12125	0.007317	0.013533	0.000333	0.0001	0	0.03175	0.001	0.0042	0.001567	0.003433	36.14	clear-nig
14	1.45E+09	1.459867	0.00345	1.459867	5.00E-05	0.021883	0.577467	0.43995	0.121033	0.007233	0.013517	0.000367	8.33E-05	1.67E-05	0.031783	0.001	0.004217	0.001583	0.00345	36.14	clear-nig
15	1.45E+09	0.840583	0.003433	0.840583	0	0.02095	0.1448	0.444783	0.035017	0.007033	0.013183	0.00065	0.000183	1.67E-05	0.031783	0.001017	0.004217	0.001617	0.003433	36.14	clear-nig
16	1.45E+09	0.7032	0.003433	0.7032	1.67E-05	0.020733	0.061967	0.443833	0.004783	0.006967	0.013117	0.000733	0.000233	0	0.03175	0.001017	0.004233	0.00155	0.003433	36.14	clear-nig
17	1.45E+09	0.571883	0.00345	0.571883	0	0.02065	0.06365	0.307783	0.004917	0.00705	0.0131	0.000733	0.00015	0	0.031733	0.001	0.004217	0.001583	0.00345	36.14	clear-nig
18	1.45E+09	0.485733	0.00345	0.485733	1.67E-05	0.020617	0.063433	0.22045	0.004983	0.007033	0.013117	0.00075	8.33E-05	0	0.031833	0.001	0.004233	0.001617	0.00345	36.14	clear-nig
19	1.45E+09	0.523167	0.003433	0.523167	0	0.020633	0.062117	0.26005	0.00495	0.007	0.013083	0.000733	0.0001	1.67E-05	0.03185	0.001017	0.00425	0.001717	0.003433	36.14	clear-nig
20	1.45E+09	0.5362	0.00345	0.5362	0	0.020683	0.062917	0.272067	0.00495	0.007033	0.01315	0.000733	0.000117	0	0.031867	0.001017	0.004233	0.00165	0.00345	36.14	clear-nig
21	1.45E+09	0.53415	0.00345	0.53415	1.67E-05	0.020667	0.06265	0.270067	0.00495	0.0071	0.01315	0.000733	0.0001	0	0.0319	0.001	0.004233	0.00175	0.00345	36.14	clear-nig
22	1.45E+09	0.533817	0.00345	0.533817	0	0.020633	0.062967	0.270033	0.00495	0.007033	0.013117	0.00075	0.0001	0	0.031817	0.001017	0.004233	0.001567	0.00345	36.14	clear-nig
23	1.45E+09	0.523633	0.00345	0.523633	1.67E-05	0.02055	0.063283	0.259817	0.00495	0.007033	0.013083	0.000733	0.000117	0	0.031733	0.001017	0.004217	0.001667	0.00345	36.14	clear-nig
24	1.45E+09	0.57725	0.003417	0.57725	0	0.020683	0.10975	0.257	0.004983	0.007067	0.013133	0.000733	8.33E-05	0	0.031683	0.001017	0.004233	0.001517	0.003417	36.14	clear-nig
25	1.45E+09	0.679567	0.003433	0.679567	1.67E-05	0.020867	0.194083	0.2571	0.005017	0.007117	0.0132	0.0007	6.67E-05	1.67E-05	0.031717	0.001017	0.004217	0.00165	0.003433	36.14	clear-nig

Figure 2: Original dataset in .csv file

The figure 2 is showing the dataset, which is taken from the kaggle machine learning website.

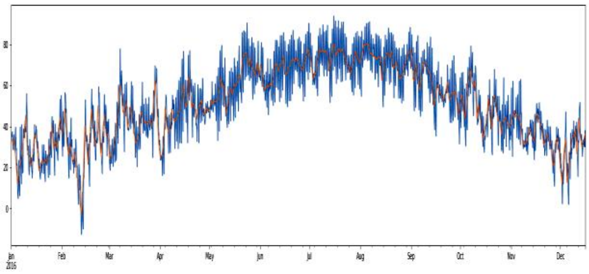


Figure 3: IOT device with month

Figure 3 is presenting various IOT devices total count with the month format. The month consider from the jan to dec.

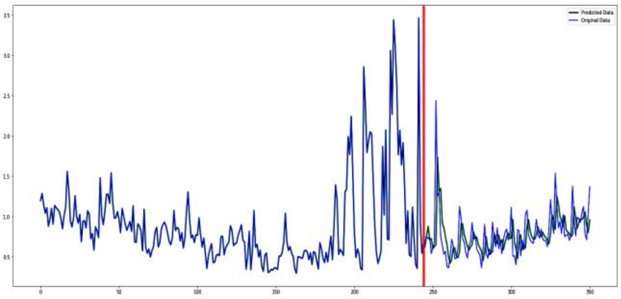


Figure 4: Predicted and actual data

Figure is 4 is showing predicted and actual data from the given dataset. It is clear from the result graph; the most of the data is predicted.

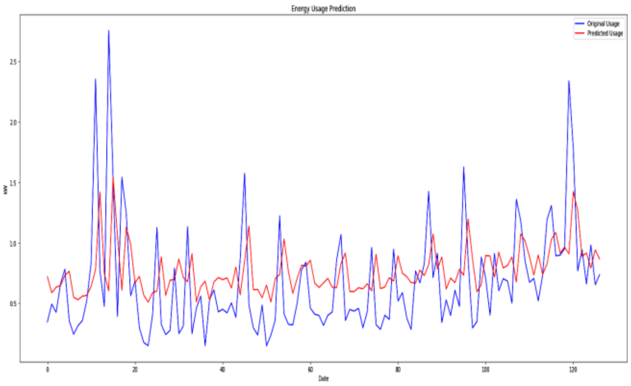


Figure 5: Energy Prediction

Figure is 5 is showing energy prediction. The original and predicted energy is shown in the graphical form.

Table 1: Result Comparison

Sr. No.	Parameters	Previous Work [1]	Proposed Work
1	Technique	Machine Learning (Linear Model)	Deep Learning (RNN-LSTM)
2	Accuracy	91.8%	95.72 %
3	Classification Error	8.2 %	4.28 %

V. CONCLUSION

This paper presents an efficient spam detection technique for IOT devices using deep learning technique. The simulation is performed using Python spyder environment, simulated results shows that the overall accuracy achieved by the proposed work is 95.72 % while previous it is achieved 91.8 %. The error rate of proposed technique is 4.28 % while 8.2 % in existing work. Therefore it is clear from the simulation results; the proposed work is achieved significant better results than existing work.

REFERENCES

1. A. Makkar, S. Garg, N. Kumar, M. S. Hossain, A. Ghoneim and M. Alrashoud, "An Efficient Spam Detection Technique for IoT Devices Using Machine Learning," in IEEE Transactions on Industrial Informatics, vol. 17, no. 2, pp. 903-912, Feb. 2021, doi: 10.1109/TII.2020.2968927.
2. F. Hossain, M. N. Uddin and R. K. Halder, "Analysis of Optimized Machine Learning and Deep Learning Techniques for Spam Detection," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), 2021, pp. 1-7.
3. A. Makkar, U. Ghosh, P. K. Sharma and A. Javed, "A Fuzzy-based approach to Enhance Cyber Defence Security for Next-generation IoT," in IEEE Internet of Things Journal, doi: 10.1109/JIOT.2021.3053326.
4. G. Fortino, F. Messina, D. Rosaci and G. M. L. Sarne, "ResIoT: An IoT social framework resilient to malicious activities," in IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 5, pp. 1263-1278, September 2020, doi: 10.1109/JAS.2020.1003330.
5. K. A. Al-Thelaya, T. S. Al-Nethary and E. Y. Ramadan, "Social Networks Spam Detection Using Graph-Based Features Analysis and Sequence of Interactions Between Users," 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT), 2020, pp. 206-211, doi: 10.1109/ICIoT48696.2020.9089509.
6. T. Y. Ho, W. Chen, M. Sun and C. Huang, "Visualizing the Malicious of Your Network Traffic by Explained Deep Learning," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 687-692.
7. J. Zhang, Q. Yan and M. Wang, "Evasion Attacks Based on Wasserstein Generative Adversarial Network," 2019 Computing, Communications and IoT Applications (ComComAp), 2019, pp. 454-459.
8. A. Makkar, N. Kumar and M. Guizani, "The Power of AI in IoT : Cognitive IoT-based Scheme for Web Spam Detection," 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019, pp. 3132-3138.
9. A. K. Singh, S. Bhushan and S. Vij, "Filtering spam messages and mails using fuzzy C means algorithm," 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), 2019, pp. 1-5, doi: 10.1109/IoT-SIU.2019.8777483.
10. T. Lange and H. Kettani, "On Security Threats of Botnets to Cyber Systems," 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), 2019, pp. 176-183, doi: 10.1109/SPIN.2019.8711780.
11. T. Qiu, H. Wang, K. Li, H. Ning, A. K. Sangaiah and B. Chen, "SIGMM: A Novel Machine Learning Algorithm for Spammer Identification in Industrial Mobile Cloud Computing," in IEEE Transactions on Industrial Informatics, vol. 15, no. 4, pp. 2349-2359, April 2019, doi: 10.1109/TII.2018.2799907.
12. G. Kumar and V. Rishiwal, "Statistical Analysis of Tweeter Data Using Language Model With KLD," 2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU), 2018, pp. 1-6, doi: 10.1109/IoT-SIU.2018.8519938