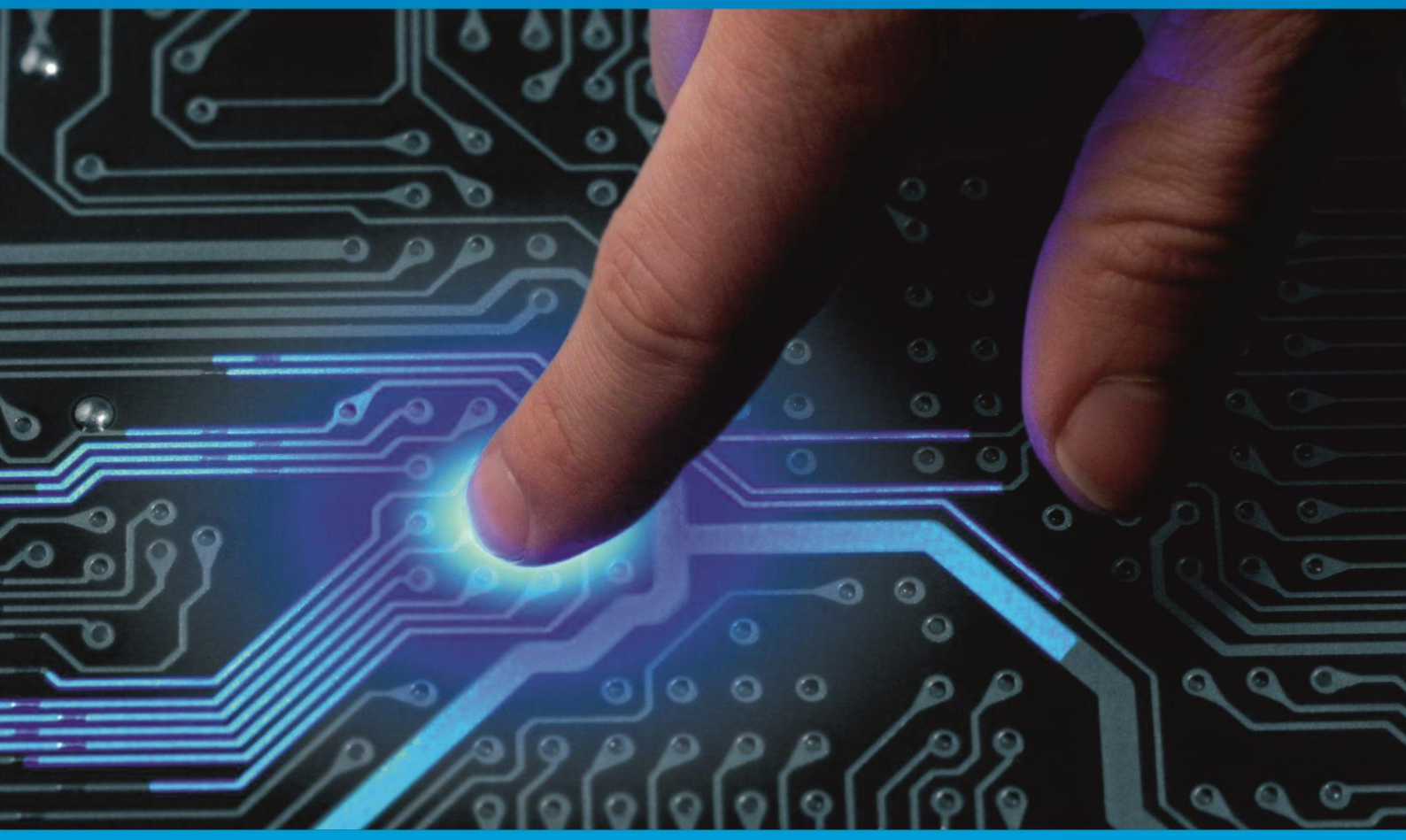
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 1, January 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0901011|

Network Health Intelligence Using Monitoring Tools and Machine Learning Algorithms

Harishchandra Amrale, Prof. Dr.Pankaj Agarkar

Department of Computer Engineering, DYPSOE, Lohegaon Pune, Savitribai Phule Pune University Pune, India

ABSTRACT: Computer networks form the basic building block of any organization's IT infrastructure. As businesses depending mostly on Internet-based applications, it is very crucial that the end users remain functioning in spite of network related issues. Monitoring and improved network performance is a high priority to keep businesses running with high availability. This assures service level agreements (SLA) maintained and provide reliable solutions to complex business problems. The Network is error prone and affects overall business performance. Therefore to maintain reliability and availability of business critical applications proactive monitoring and actions based on the monitoring have uttermost importance. Proactive monitoring identifies the issues and trigger the corrective actions before it actually experiences by the users. Network Monitoring is the one part of our proposed solution where already many Open source tools like Nagios and Zabbix can be part of the solution. The Second part of the proposed solution is relying on data collection by network monitoring tools build the analytical based solution, which can be predict the health of network component in advance. Our proposed Network health intelligence system not only efficiently collects the monitoring data but also using machine learning algorithms predict the health of network components inadvance.

KEYWORDS: Supervised machine learning models, unsupervised machine learning models, Messaging systems, and Performance metrics

I.INTRODUCTION

As Computer networks form the basic building block of any organization's IT infrastructure. As businesses depending mostly on Internet-based applications, it is very crucial thatthe end users remain functioning in spite of network related issues. Monitoring and improved network performance is high priority to keep businesses running with high availability. This assures service level agreements (SLA) maintained and provide reliable solutions to complex business problems.The Network is error prone and affects overall business performance. Therefore to maintain reliability and availability of business critical applications proactive monitoring and actions based on the monitoring have uttermost importance. Proactive monitoring identifies the issues and triggers the corrective actions before it actually experience by the users. Proactive monitoring is the one step but predictive analysis based on the previous performance history of thenetwork

II.LITERATURE SURVEY

RaoufBoutaba, Mohammad Ali Salahuddin, NouraLimam, Sara Ayoubi, NashidShahriar, Felipe Estrada Solano, Oscar Mauricio CaicedoRendon: A comprehensive survey on machine learning for networking evolution, applications and research opportunities. In this paper authors provide the survey on the application of ML techniques in a variety of network areas and network technologies. It provides a comprehensive analysis of Machine learning application to basic issues in networking like traffic prediction, traffic routing, and fault management and networksecurity.[1]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0901011|

1. Rafiullah Khan, SarmadUllah Khan, RifaqatZaheer, and Muhammad Inayatullah Babar: An Efficient Network Monitoring and Management System This paper discussed the continuous and automated way to monitor the network components i.e. Network switches. It provides the scope for triggering mechanisms like SMS, EMAIL to the network administrator. It also proposes an automated way of ticket creation and assignment.[2]

2. Kwang-Bon Jung,Mi-Jung Choi,Myung-Sup Kim,Young-J. Won,James W. Hong: Traffic Classification Using Machine Learning Algorithms in Practical Network Monitoring Environments. As per this paper, the network traffic classification went through an evolution from payload/port based to machine learning based. This paper analyzes the classification results from cross-validation with split validation. This paper also discussed the classification results depends on flow to those depends on bytes. Authors provide the comparative report of different machine learning algorithms like J48, REPTree, RBFNetwork, Multilayer perceptron, BayesNet, and NaiveBayes. In this paper, Author identifies the best feature sets and the best ML algorithm for network traffic classification using the splitvalidation.[3]

3. Mowei Wang, Yong Cui, Xin Wang, Shihan Xiao, and Junchen Jiang:Machine Learning for Networking: Workflow, Advances and Opportunities. This paper discussed the basic overflow of the application of machine learning algorithms in the networking [4]

4. Deepak Chahal, LatikaKharb, DeepanshuChoudhary: Performance Analytics of Network Monitoring Tools. TheNetwork is the backbone of the organization's infrastructure and maintaining its availability and reliability is always the challenge. In this paper, authors discussed some popular network monitoring tools like Nagios,Zabbix and ganglia and provide the comparison report of different network monitoring tools based on the parameters like license, access control and distributedmonitoring.[5]

5. SaacSikubwabo, Mariam Usanase, Dr. PapiasNiyigena:Comparative Study on Network Monitoring Tools of NagiosVersusHyperic.Author provides the comparison between two well known network monitoring tools NagiosandHyperic[6]

III.PROPOSED METHODOLOGY

Network components performance metrics collection using network monitoring tool and identification, Implementation of supervised and unsupervised machine learning models is the overall scope of the project. Although scope is not limited to the application of existing machine learning algorithms but also improvement in existing machine learning techniques/algorithms with developing new interfaces to available network monitoring tools also the part ofthe scope.

First, we will browse nagios and collect data into CSV format .The data get into csv format will be store into MySQL Database.There are Three modules that are Resource utilization, Network and Threat. Before this,we will perform preprocessing

Data Pre-processing is a technique that is used to convert the raw data into a clean data set.

The dataset available to us in a raw form, i.e. the title, content and the tags in the training dataset is unstructured data. We then process this unstructured data.

- **Data pre-processing includes**:

1. Cleaning,
2. Instance selection,
3. Normalization,
4. Transformation,
5. Feature extraction and
6. Selection

The product of **data pre-processing** is the final training set.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0901011|

- **Cleaning:**

Data cleaning is the process of fill in missing values, smoothing the noisy data, identify or remove outliers, and resolve inconsistencies.

- **Instance selection:**

The process of ensuring that your data is correct, consistent and useable by identifying any errors or corruptions in the data, correcting or deleting them, or manually processing them as needed to prevent the error from happening again. Instance selection should identify a subset of the total available data to achieve the original purpose of the data mining (or machine learning) application as if the whole data had been used. Considering this, the optimal outcome of IS would be the minimum data subset that can accomplish the same task with no performance loss, in comparison with the performance achieved when the task is performed using the whole available data. Therefore, every instance selection strategy should deal with a trade-off between the reduction rate of the dataset and the classification quality.

- **Normalization:**

Data normalization is a pre-processing technique where it groups the given data into a well refined format. The data normalization technique is used to manage and organize the feature values in the dataset. Also, it scales the feature values to the same specified range.

- **Transformation:**

Data transformation is the process of converting data or information from one format to another, usually from the format of a source system into the required format of a new destination system.

- **Feature extraction:**

Feature extraction is used here to identify key features in the data for coding by learning from the coding of the original data set to derive new ones. – A technique for natural language processing that extracts the words (features) used in a sentence, document, website, etc. and classifies them by frequency of use.

Data Preprocessing for Machine Learning

1. Step 1: Import Libraries. The First step is usually importing the libraries that will be needed in the program.
2. Step 2: Import the Dataset.
3. Step 3: Taking care of Missing Data in Dataset.
4. Step 4: Encoding categorical data.
5. Step 5: Splitting the Dataset into the Training set and Test Set.
6. Step 6: Feature Scaling.

For Classification into three modules i.e. Resource, Network and Threat we will use Naïve Bayes Algorithm.
Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier.
This data will be classified into these three modules that is Resource, Network and Threat.

- **Resource Utilization**

Inside resource utilization, we need to predict the failure (Resource utilization cross the threshold limit) that will occur in a month. The parameters inside resource for prediction is CPU, RAM, and HARD DISK. These the parameters will be a check on the usage and one threshold limit will be assigned for these three parameters and will be predicted

- **Network**

Inside network, we need to predict the failure that will occur or can occur. The network can be used for analyzing traffic and to check we the network will be gone down. So in network, there are two parameters they are network traffic and network down (also called as network outage) need to predicate.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0901011|

- **Threat**

Inside threat, the word threat is related to security. Threat can come through a network when illegal activity is being done by an intruder or by a malicious user. The parameters come under for security threat the User Behaviors.

For prediction SVM (Support Vector Machine) algorithm is used for predicting the further failure that will be occurred based on parameters under each module that is Resource Prediction, Network Prediction and Security Prediction

A support vector machine takes these data points and outputs the hyperplane (which in two dimensions it's simply a line) that best separates the tags. This line is the decision boundary: anything that falls to one side of it we will classify as blue, and anything that falls to the other as red.

A. Architecture

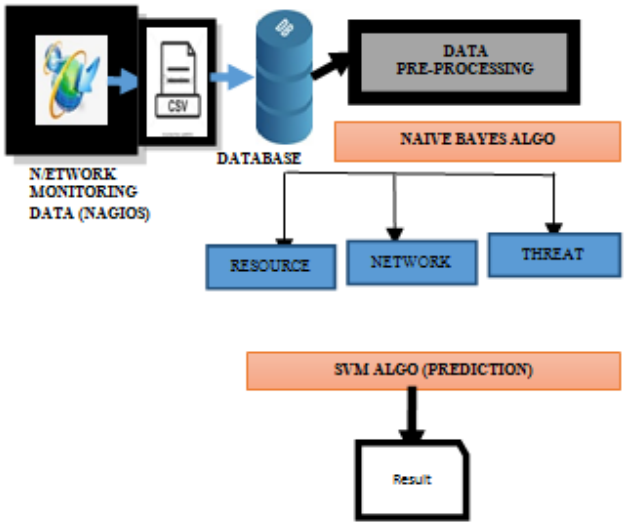


Fig. 1. Proposed System Architecture

Algorithm

Naive-Bayes Classification

- Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group/class. In short, it is a probabilistic classifier.

- The Naive Bayes algorithm is called "naive" because it assumes that the occurrence of a certain feature is independent of the occurrence of other features.

The Mathematics of the Naive Bayes Algorithm

The basis of the Naive Bayes algorithm is Bayes' theorem or alternatively known as Bayes' rule or Bayes' law. It gives us a method to calculate the conditional probability, i.e., the probability of an event based on previous knowledge available on the events. More formally, Bayes' Theorem is stated as the following equation:

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0901011|

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Let us understand the statement first and then we will look at the proof of the statement. The components of the above statement are:

- P (A/B): Probability (conditional probability) of occurrence of event A given the event B is true
- P(A) and P(B): Probabilities of the occurrence of event A and B respectively
- P(B/A): Probability of the occurrence of event B given the event A is true

The terminology in the Bayesian method of probability (more commonly used) is as follows:

- A is called the proposition and B is called the evidence.
- P(A) is called the prior probability of proposition and P(B) is called the prior probability of evidence.
- P (A/B) is called the posterior.
- P (B/A) is the likelihood.

This sums the Bayes' theorem as

$$\text{Posterior} = \frac{(\text{Likelihood}).(\text{Proposition prior probability})}{\text{Evidence prior probability}}$$

2. Support Vector Machine (SVM) Algorithm:

A support vector machine takes these data points and outputs the hyperplane (which in two dimensions it's simply a line) that best separates the tags. This line is the decision boundary: anything that falls to one side of it we will classify as **blue**, and anything that falls to the other as red.

The goal of SVM is to divide the datasets into classes to find a maximum marginal hyperplane

Algorithm Steps:

To create the SVM classifier, we will import SVC class from Sklearn. svm library.

1. from sklearn. svm import SVC # "Support vector classifier"
2. classifier = SVC(kernel='linear', random_state=0)
3. classifier. fit(x_train, y_train)
4. Import the dataset
5. Explore the data to figure out what they look like
6. Pre-process the data
7. Split the data into attributes and labels
8. Divide the data into training and testing sets
9. Train the SVM algorithm
10. Predictions

Pseudocode:

```
Importjava.io.File;
importjava.io.FileNotFoundException;
```

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0901011|

```
importjava.io.FileOutputStream;
importjava.io.IOException;
importjava.io.OutputStreamWriter;
importjava.io.Writer;
/*** Represent the classification function* @author defaultstr**/
public class Classifier {
private Kernel k;
public double[] alpha;
publicDataPoint[] sv;
public double b;
/** classification function* @param k kernel function
* @param alpha multiplier of support vectors
* @paramsv support vectors
* @param b b in classification function*/
public Classifier(Kernel k,
double[] alpha,
DataPoint[] sv,
double b) {
this.k = k;
this.alpha = alpha;
this.sv = sv;
this.b = b;
}
/**
* predict which class the input DataPoint belongs to
* @param p input DataPoint
* @return the class y predicted by SVM, y in {+1,-1}
*/
publicint classify(DataPoint p) {
double f = 0.0;
try {
for (int i = 0; i <alpha.length; i++) {
f += alpha[i]*sv[i].y*k.k(p, sv[i]);
}
} catch (DataPointTypeMismatchException e) {
e.printStackTrace();
}
f += b;
if (f >= 0)
return 1;
else
return -1;
}
```

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0901011|

```
/**
* classify all the DataPoint in input array
* the result will be written in DataPoint's member y
* @param p
*/
public void classifyAll(DataPoint[] p) {
for (DataPointdp : p) {
dp.y = classify(dp);
}
}
public void output2DSupportVector(String filename) throws IOException {
// for matlab
File fout = new File(filename);
Writer w = new OutputStreamWriter(new FileOutputStream(fout));
w.write(b + "\n");
SparseVectorPointvp;
for (int i = 0; i <sv.length; i++) {
vp = (SparseVectorPoint)sv[i];
w.write(sv[i].y + " " + alpha[i] + " " + vp.val[0] + " " + vp.val[1] + "\n");
}
w.close();
}
}
```

IV.RESULT AND DISCUSSION

Experiments are done by a personal computer with a configuration: Intel (R) Core (TM) i3-2120 CPU @ 3.30GHz, 4GB memory, Windows 7, MySQL 5.1 backend database and jdk 1.8. The application is dynamic web application for design code in Eclipse tool and execute on Tomcat server. Some functions used in the algorithm are provided by list of jars like standford core NLP jar for keywords extraction using tagger method.

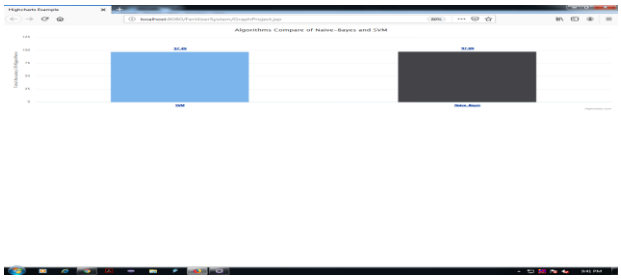
Number	Algorithms	Value (%)
1.	SVM	97.3
2.	Naive-Bayes Classification	97.6

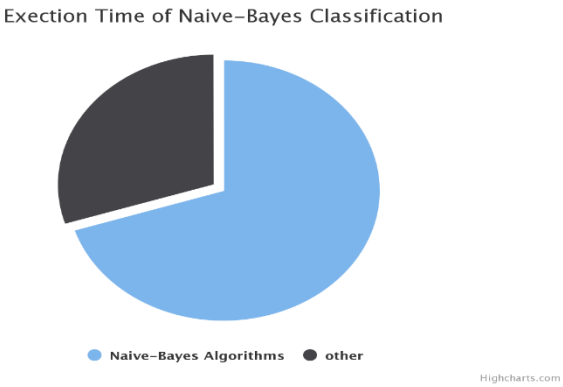
Table 1 Compare Table of Algorithms

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 1, January 2021 ||

| DOI: 10.15680/IJIRCCE.2021. 0901011|





V.CONCLUSION

The Network comprises of different complex components and becomes more error prone with the addition of each new component. Overall it is required of early detection of network errors and necessary actions to remove the errors before it become visible to the users. Proactive monitoring is the one step but predictive analysis based on the previous performance history of the network components provides the edge. Network components performance metrics collection using network monitoring tool and identification, Implementation of supervised and unsupervised machine learning models is the overall scope of the project. Although the scope is not limited to the application of existing machine learning algorithms but enhancement of machine learning algorithms with developing new interfaces to available network monitoring tools also the part of the scope. Storage strategy to store performance metrics collected by network monitoring tool is an essential part of overall scope e.g. Cluster formation/Big data setup.

REFERENCES

[1] RaoufBoutaba, Mohammad Ali Salahuddin, NouraLimam, Sara Ayoubi, NashidShahriar, Felipe Estrada Solano, Oscar Mauricio CaicedoRendon"A comprehensive survey on machine learning for networking evolution, applications and research opportunities,"Journal of Internet services and applications,2018

[2] Rafiullah Khan, SarmadUllah Khan, RifaqatZaheer, and Muhammad Inayatullah Babar "An Efficient Network Monitoring and Management System,"International Journal of Information and Electronics Engineering,2013

[3] Kwang-Bon Jung,Mi-Jung Choi,Myung-Sup Kim,Young-J. Won,James,January 2008, The journal of Korean Institute of communications and information sciences

[4] Mowei Wang, Yong Cui, Xin Wang, Shihan Xiao, and Junchen Jiang "Machine Learning for NetworkingWorkflow, Advances and Opportunities," IEE Network ,2017

[5] Deepak Chahal, LatikaKharb, DeepanshuChoudhary"Performance Analytics of Network Monitoring Tools,"IJITEE ,Volume 8,Issue8,2019

[6] SaacSikubwabo, Mariam Usanase, Dr. PapiasNiyigena"Comparative Study on Network Monitoring Tools of Nagios Versus Hyperic, "IRJET,Volume 6,Issue12,2019

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor:
7.488

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details