Hello, this is Bing. I’m happy to help you with your query.😊

[According to the search results, **Accurate Feature Elimination approach in Ensemble Learning on NSL-KDD dataset** is a method of selecting the most relevant features for network anomaly detection by using an ensemble of different feature selection techniques and comparing them with the Candidate Elimination Algorithm1](https://ieeexplore.ieee.org/document/10179832/). [This method claims to achieve higher accuracy and lower false positive rate than the Candidate Elimination Algorithm1](https://ieeexplore.ieee.org/document/10179832/).

[Some of the **roles and responsibilities** of a data scientist are2](https://www.simplilearn.com/data-scientist-job-description-article)[3](https://www.geeksforgeeks.org/what-are-the-roles-and-responsibilities-of-a-data-scientist/)[4](https://www.indeed.com/hire/job-description/data-scientist):

* Collecting data from various sources and analyzing it to find patterns, trends and insights
* Using statistical and machine learning tools to create and optimize models, classifiers, algorithms and solutions for various business problems
* Enhancing data collection and preprocessing procedures to ensure data quality and reliability
* Communicating and presenting the results and recommendations to stakeholders in a clear and compelling way
* Collaborating with other data scientists, engineers, analysts and business teams to deliver data-driven products and services

To explain these roles and responsibilities in interviews, you should be able to demonstrate your skills and experience in:

* Data mining, wrangling, cleaning, manipulation and visualization
* Programming languages such as Python, R, SQL, etc.
* Machine learning frameworks such as TensorFlow, PyTorch, Scikit-learn, etc.
* Data analysis techniques such as regression, classification, clustering, anomaly detection, etc.
* Data storytelling and reporting tools such as Tableau, Power BI, etc.

You should also be able to showcase your projects and portfolio that highlight your data science capabilities and achievements. You should be prepared to explain the business problem, the data sources, the methods used, the results obtained and the impact or value created by your solution.

I hope this helps you prepare for your interviews. Good luck!👍

[The **NSL-KDD dataset** is a data set for network-based intrusion detection systems (IDS) that was proposed to solve some of the inherent problems of the original KDD’99 dataset1](https://www.kaggle.com/datasets/hassan06/nslkdd)[2](https://www.impactcybertrust.org/dataset_view?idDataset=928). Some of the problems of the KDD’99 dataset are:

* It contains a large number of redundant records, which makes the learning algorithms biased towards more frequent records and prevents them from learning rare but harmful attacks
* It has duplicate records in the test set, which makes the evaluation results unreliable and inconsistent
* It has an imbalanced distribution of records from different difficulty levels, which makes it hard to compare the performance of different learning techniques

[The NSL-KDD dataset has the following advantages over the KDD’99 dataset3](https://www.unb.ca/cic/datasets/nsl.html)[2](https://www.impactcybertrust.org/dataset_view?idDataset=928):

* It does not include redundant records in the train set, so the classifiers are not biased towards more frequent records
* It does not have duplicate records in the test sets, so the performance of the learners are not biased by the methods that have better detection rates on the frequent records
* It has a proportional selection of records from each difficulty level group, which makes it more efficient to have an accurate evaluation of different learning techniques
* It has a reasonable number of records in the train and test sets, which makes it affordable to run the experiments on the complete set without the need to randomly select a small portion

The NSL-KDD dataset contains 41 features and 125000 samples, and it is divided into two subsets: KDDTrain+ and KDDTest+. The KDDTrain+ subset contains 67% of the selected records and it is used for training purposes. The KDDTest+ subset contains 33% of the selected records and it is used for testing purposes. The dataset also provides attack-type labels and difficulty levels for each record. [The attack types are categorized into four main groups: Normal, DoS, Probe, R2L, and U2R3](https://www.unb.ca/cic/datasets/nsl.html)[1](https://www.kaggle.com/datasets/hassan06/nslkdd).

You can download the NSL-KDD dataset from these links:

* https://www.unb.ca/cic/datasets/nsl.html
* https://www.kaggle.com/datasets/hassan06/

[**NSL-KDD**](https://www.bing.com/ck/a?!&&p=4180a2f86f1cf40dJmltdHM9MTY5MDc2MTYwMCZpZ3VpZD0wNDAzMmY1Yi0zM2I2LTYxMmItMGU5Ni0zZmZiMzIxYjYwMmYmaW5zaWQ9OTQzNQ&ptn=3&hsh=3&fclid=04032f5b-33b6-612b-0e96-3ffb321b602f&psq=NSL-KDD+dataset&u=a1aHR0cHM6Ly93d3cudW5iLmNhL2NpYy9kYXRhc2V0cy9uc2wuaHRtbA&ntb=1)

https://www.unb.ca/cic/datasets/nsl.html

* **The NSL-KDD data set has the following advantages over the original KDD data set:**
  + It does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records.
  + There is no duplicate records in the proposed test sets; therefore, the performance of the learners are not biased by the methods which have better detection rates on the frequent ...
  + The number of selected records from each difficultylevel group is inversely proportional to the percentage of records in the original KDD data set. ...

Tell me about your self

Hi sir/mam good evening

I am N Kondaiah Naidu and I am from Tirupathi.

As I completed my BE graduation in Saveetha School Of Engineering in Chennai in the main stream of Computer Science And Engineering with aggregate percentage of 72%.

And I completed my +2 intermediate in Narayana Junior College Tirupathi With aggregate of 86%

And I completed my schooling in Sri Chaitanya School Tirupathi With aggregate of 88% This Is about my studying.

I am belongs in a small family like me and my father, mother and one elder sister

My father is Mr. N Venkata Ramana and he is Agriculturalist.

And my mother Mrs. N Chandra Kala and she is Home maker.

And I have elder sister Ms. N Divya Sree and is working in 24/7 infotech in Bangalore.

My strengths are I am hardworking and self motivated person towards my carrer and my life

My weekness is I cant say no when any one ask for help.

My goals are Short term Goal is to get a job in reputed company where I can utilize my skills and improve my carrer path.

My long term goal is to be in respectable position in that organization.

My hobbies are Reading, gardening and playing games.

That’s all about my self and thanks for giving this golden opportunity to introduce my self sir.

About Project:

I have chosen the Domain as machine learning and I have taken the one Main title and I have taken 4 comparison algorithms.

* 1. Ensemble Learning in comparison with candidate Elimination algorithm.
  2. Ensemble Learning in comparison with Variable Elimination algorithm.
  3. Ensemble Learning in comparison with List-Then Elimination algorithm.
  4. Ensemble Learning in comparison with Version Space Elimination algorithm.

And I have downloaded the NSL-KDD dataset from Kaggle.com and in this NSL means Network Security Language and KDD means Knowledge Discovery in Data bases.

The KDD Dataset is a dataset for Network-intrusion detection system that was proposed to solve some of the inherient problems of the original KDD dataset

I have taken the IBM SPSS software to calculate result in statistics manner to show the proposed algorithm is better compared to the existing algorithm.

With the help of the 4 comparison algorithms I will compare with the main algorithm and I prove that the proposed algorithm is providing better accuracy and predictive results.

What do you know about Accenture in interview?

About Accenture  
  
Accenture is an Irish multinational company that offers professional services in the field of Information technology and consulting. And that helps the world's leading businesses, governments and other organizations build their digital core, optimize their operations, accelerate revenue growth and enhance citizen services—creating tangible value at speed and scale

Why should I hire you?

Eventhough I am a fresher I have theoriotical Knowledge but can do hardwork for my organization and I will put all my efforts for good progress in organization

Hi mam good morning

I am n kondaiah naidu and I am from tirupathi.

As I completed my be graduation in saveetha school of engineering Chennai in the main stream of computer science and engineering and I passed in distinction.

My short term goal is to get a job in reputed company where I can utilize my skills and improve my carrier path

My long term goal is to be a respectable position in that organization by learning new technologies which is required.

My strengths are I am hardworking person and I have to become an enterperneur in my life.

My weakness are as I born in a middle class family first I have to stand in my life and I can fulfil my goal as a enterpernuer

My hobbies are Playing cricket, gardening and browsing the internet.

Any questions

May I know what achievements u have made and do u faced any problems in your tenure I am very curious to know these things.

Project:

Title: Accurate Feature Elimination Approach in Ensemble Learning on NSL-KDD Dataset.

In this NSL means Network Security Language

And KDD means Knowledge Discovery In Databases

The KDD dataset is a dataset for Network Intrusion Detection system that was proposed to solve some of inherient problems present in the original KDD dataset.

Novel **Feature elimination** for **NSL**-**KDD dataset** is collected from kaggle.com, this dataset is applied to. **Ensemble Learning** and Candidate Elimination Algorithm.

Advantages:

1.Feature elimination offers a simple effective way to overcome this challenge by eliminating redundant and irrelevant data.

2.Removing of irrelevant data improves accuracy of results.

3.Reduces the computation time, and facilitates an enhanced understanding of the model or data.

4. It does not include redundant records in the train set, so the classifiers will not be biased towards more frequent records.

Dis Advantages:

What are the drawbacks of NSL-KDD dataset?

One of the most important disadvantage in the KDD data set is the huge number of redundant records, which causes the learning algorithms to be biased towards the frequent records, and thus prevent them from learning unfrequent records which are usually more harmful to networks such as U2R and R2L attacks.