**MODELS PERFORMANCE WITH COMPARISON:**

**Naïve Bayes:**

| **Attacks** | **15** | **24** | **48** | **All** |
| --- | --- | --- | --- | --- |
| **Bot(0)** | **0.00000** | **0.00000** | **0.09135** | **0.08669** |
| **Brute Force(1)** | **0.00000** | **0.00000** | **0.00692** | **0.04234** |
| **Dos/DdoS(2)** | **0.45384** | **0.45900** | **0.44249** | **0.49844** |
| **Infiltration(3)** | **0.00000** | **0.45900** | **0.00000** | **0.00873** |
| **Normal(4)** | **0.84343** | **0.84763** | **0.75491** | **0.76480** |
| **PortScan(5)** | **0.97945** | **0.97945** | **0.97138** | **0.97367** |
| **Web Attack(6)** | **0.15807** | **0.14774** | **0.01486** | **0.01530** |
| **Accuracy** | **0.76194** | **0.76726** | **0.65155** | **0.64297** |
| **FAR** | **0.23806** | **0.23274** | **0.34845** | **0.35703** |

**Winner is 24**

**For Naïve Bayes model 15(0.76194) and 24(0.76726) features performance is similar and 48(0.65155) and All(0.64297) features performance is similar. If you compare the different features in the table you will find that 24 features are performing well. If you give more importance to the no of features then you can go for 15 features because the accuracy values of 15 and 24 features are similar.**

**KNN:**

| **Attacks** | **15** | **24** | **48** | **All** |
| --- | --- | --- | --- | --- |
| **Bot(0)** | **0.84762** | **0.83810** | **0.86239** | **0.86239** |
| **Brute Force(1)** | **0.96757** | **0.97892** | **0.98165** | **0.97900** |
| **Dos/DdoS(2)** | **0.99022** | **0.99244** | **0.99200** | **0.99291** |
| **Infiltration(3)** | **0.00000** | **0.00000** | **0.00000** | **0.00000** |
| **Normal(4)** | **0.99618** | **0.99672** | **0.99680** | **0.99691** |
| **PortScan(5)** | **0.95613** | **0.96081** | **0.96320** | **0.96347** |
| **Web Attack(6)** | **0.67586** | **0.70307** | **0.72368** | **0.71895** |
| **Accuracy** | **0.99354** | **0.99448** | **0.99461** | **0.99481** |
| **FAR** | **0.00646** | **0.00552** | **0.00539** | **0.00519** |

**Winner is ALL**

**For KNN model 24(0.99448), 48(0.99461) and All(0.99481) features are performing similarly. If you give more importance to no of features you can choose the 24 features, because it is performing well with less features. Overall the ALL features performing well.**

**RF:**

| **Attacks** | **15** | **24** | **48** | **All** |
| --- | --- | --- | --- | --- |
| **Bot(0)** | **0.95775** | **0.96226** | **0.87437** | **0.80423** |
| **Brute Force(1)** | **0.99818** | **0.99818** | **0.99635** | **0.99635** |
| **Dos/DdoS(2)** | **0.99935** | **0.99914** | **0.99848** | **0.99814** |
| **Infiltration(3)** | **0.00000** | **0.00000** | **0.00000** | **0.00000** |
| **Normal(4)** | **0.99966** | **0.99980** | **0.99964** | **0.99954** |
| **PortScan(5)** | **0.99982** | **0.99991** | **0.99973** | **0.99973** |
| **Web Attack(6)** | **0.82734** | **0.98529** | **0.98529** | **0.98529** |
| **Accuracy** | **0.99943** | **0.99967** | **0.99939** | **0.99922** |
| **FAR** | **0.00057** | **0.00033** | **0.00061** | **0.00078** |

**Winner is 24**

**For Random Forest 24 features is performing well with accuracy 0.99967. If you observe the above table you can see that 15 features(0.99943) and 48 features(0.99939) performance is similar. False alarm rate of 24 features(0.00033) is much lower when comparatively to other features.**

**XGB:**

| **Attacks** | **15** | **24** | **48** | **All** |
| --- | --- | --- | --- | --- |
| **Bot(0)** | **0.98165** | **0.97696** | **0.98165** | **0.97696** |
| **Brute Force(1)** | **0.99909** | **0.99818** | **0.99818** | **1.00000** |
| **Dos/DdoS(2)** | **0.99950** | **0.99942** | **0.99958** | **0.99956** |
| **Infiltration(3)** | **0.50000** | **0.50000** | **0.00000** | **0.00000** |
| **Normal(4)** | **0.99979** | **0.99986** | **0.99989** | **0.99989** |
| **PortScan(5)** | **1.00000** | **0.99991** | **1.00000** | **1.00000** |
| **Web Attack(6)** | **0.89209** | **0.98529** | **0.98529** | **0.98529** |
| **Accuracy** | **0.99964** | **0.99977** | **0.99981** | **0.99981** |
| **FAR** | **0.00036** | **0.00023** | **0.00019** | **0.00019** |

**Winner is both 48 and ALL**

**For XGB 48 and All features are performing equally. So, you can choose anyone one of them. There is slight difference in the f1 scores of the 48 and All features, but overall accuracy is similar i.e 0.99981. if you give more importance to the no of features then you can choose the 48 features.**

**If you compare the 24(0.99977) and 48(0.99981) features there performance is slightly similar. If you give more preference to less features we need to trade off the performance, but this trade off does not effect the performance to greater extent. So, in that case we can choose 24 features.**

**If you see the 15 features it is performing well even though it is having less no of features. If you choose for the lowest features then 15 features is the best for you.**

**DCT:**

| **Attacks** | **15** | **24** | **48** | **All** |
| --- | --- | --- | --- | --- |
| **Bot(0)** | **0.90741** | **0.92523** | **0.93897** | **0.92958** |
| **Brute Force(1)** | **0.99818** | **0.99547** | **0.99909** | **0.99637** |
| **Dos/DdoS(2)** | **0.99901** | **0.99888** | **0.99872** | **0.99895** |
| **Infiltration(3)** | **1.00000** | **0.00000** | **0.66667** | **0.40000** |
| **Normal(4)** | **0.99958** | **0.99969** | **0.99972** | **0.99973** |
| **PortScan(5)** | **0.99982** | **0.99963** | **0.99982** | **0.99945** |
| **Web Attack(6)** | **0.82394** | **0.93750** | **0.96085** | **0.96429** |
| **Accuracy** | **0.99929** | **0.99946** | **0.99951** | **0.99953** |
| **FAR** | **0.00071** | **0.00054** | **0.00049** | **0.00047** |

**Winner is ALL**

**For DCT 48(0.99951) and All(0.99953) features performance is similar. Overall ALL features is best for model. Further we can compare the results of DCT with Bagging, Boosting and Stacking.**

**Bagging:**

| **Attacks** | **15** | **24** | **48** | **All** |
| --- | --- | --- | --- | --- |
| **Bot(0)** | **0.96330** | **0.96744** | **0.97222** | **0.97696** |
| **Brute Force(1)** | **0.99637** | **0.99728** | **0.99728** | **0.99728** |
| **Dos/DdoS(2)** | **0.99911** | **0.99916** | **0.99903** | **0.99916** |
| **Infiltration(3)** | **0.50000** | **0.80000** | **0.80000** | **0.00000** |
| **Normal(4)** | **0.99968** | **0.99981** | **0.99980** | **0.99982** |
| **PortScan(5)** | **0.99991** | **1.00000** | **1.00000** | **1.00000** |
| **Web Attack(6)** | **0.87889** | **0.97473** | **0.98182** | **0.98901** |
| **Accuracy** | **0.99946** | **0.99968** | **0.99967** | **0.99971** |
| **FAR** | **0.00054** | **0.00032** | **0.00033** | **0.00029** |

**Winner is All features**

**For Bagging 24(0.99946), 48(0.99967) and All(0.99971) features performance is similar. If you choose for the less features, then you can go for 24 features. Overall All features is performing well comparatively with other features.**

**Stacking:**

| **Attacks** | **15** | **24** | **48** | **All** |
| --- | --- | --- | --- | --- |
| **Bot(0)** | **0.97696** | **0.97196** | **0.97222** | **0.97222** |
| **Brute Force(1)** | **0.99909** | **0.99909** | **0.99818** | **0.99818** |
| **Dos/DdoS(2)** | **0.99942** | **0.99953** | **0.99937** | **0.99950** |
| **Infiltration(3)** | **0.00000** | **0.00000** | **0.00000** | **0.00000** |
| **Normal(4)** | **0.99973** | **0.99987** | **0.99984** | **0.99986** |
| **PortScan(5)** | **0.99982** | **0.99991** | **0.99982** | **0.99982** |
| **Web Attack(6)** | **0.86713** | **0.98529** | **0.98529** | **0.98529** |
| **Accuracy** | **0.99955** | **0.99979** | **0.99974** | **0.99977** |
| **FAR** | **0.00045** | **0.00021** | **0.00026** | **0.00023** |

**Winner is 24**

**Before we used the models like Random Forest, Naïve Bayes and KNN separately. Random Forest lowest FAR is 0.00033 for 24 features, Naïve Bayes Lowest FAR is 0.23274 for 24 features and KNN lowest FAR is 0.00519 for ALL features. So, we stacked all three models and got very good results compared to individual results.**

**For above stacking model 24 features is performing well with accuracy 0.99979. performance of 24 features and ALL features are performing similarly.**

**Overall stacking improved the performance of the individual models.**

**LSTM:**

| Attacks | 15 | 24 | 48 | All |
| --- | --- | --- | --- | --- |
| Bot(0) | 0.65693 | 0.80000 | 0.74830 | 0.76562 |
| Brute Force(1) | 0.92582 | 0.98474 | 0.96070 | 0.93857 |
| Dos/DdoS(2) | 0.99080 | 0.99442 | 0.99266 | 0.98635 |
| Infiltration(3) | 0.00000 | 0.00000 | 0.00000 | 0.00000 |
| Normal(4) | 0.99598 | 0.99709 | 0.99683 | 0.99493 |
| PortScan(5) | 0.95961 | 0.96643 | 0.96882 | 0.94386 |
| Web Attack(6) | 0.00000 | 0.04545 | 0.00000 | 0.00000 |
| Accuracy | 0.99320 | 0.99512 | 0.99469 | 0.99130 |
| FAR | 0.0068 | 0.0049 | 0.0053 | 0.0087 |

**- The model achieves high accuracy across all attack categories, ranging from 99.130% to 99.512%. This suggests that the LSTM model is effective in detecting and classifying various types of attacks.**

**- The false alarm rate (FAR) is relatively low, ranging from 0.49% to 0.87%. This indicates that the model has a good ability to distinguish normal network traffic from malicious attacks.**

**- The Brute Force attack category shows consistently high performance, with accuracy ranging from 93.857% to 98.474%. This suggests that the model is particularly effective in detecting and classifying brute force attacks.**

**- The Infiltration attack category shows 0% accuracy across all feature configurations. This might indicate that the LSTM model struggles to effectively detect and classify Infiltration attacks.**

**- The Web Attack category shows low accuracy, with values ranging from 0% to 4.545% across different feature configurations. This suggests that the LSTM model has difficulty accurately detecting and classifying web attacks.**

**CNN:**

| Attacks | 15 | 24 | 48 | All |
| --- | --- | --- | --- | --- |
| Bot(0) | 0.81657 | 0.71429 | 0.84524 | 0.79070 |
| Brute Force(1) | 0.98413 | 0.98675 | 0.99037 | 0.98137 |
| Dos/DdoS(2) | 0.99555 | 0.99536 | 0.99665 | 0.99577 |
| Infiltration(3) | 0.00000 | 0.00000 | 1.00000 | 0.00000 |
| Normal(4) | 0.99684 | 0.99698 | 0.99694 | 0.99733 |
| PortScan(5) | 0.95447 | 0.96684 | 0.95518 | 0.95928 |
| Web Attack(6) | 0.07229 | 0.55738 | 0.03774 | 0.87342 |
| Accuracy | 0.99469 | 0.99495 | 0.99486 | 0.99546 |
| FAR | 0.0053 | 0.0051 | 0.0051 | 0.0045 |

**- The model achieves high accuracy across all attack categories, ranging from 99.469% to 99.546%. This suggests that the CNN model is effective in detecting and classifying various types of attacks.**

**- The false alarm rate (FAR) is consistently low, ranging from 0.45% to 0.53%. This indicates that the model has a good ability to distinguish normal network traffic from malicious attacks.**

**- The Brute Force attack category shows consistently high performance, with accuracy ranging from 98.413% to 99.037%. This suggests that the model is particularly effective in detecting and classifying brute force attacks.**

**- The Infiltration attack category has an accuracy of 100% for the 48-feature configuration. However, the accuracy is 0% for the 15 and 24-feature configurations. This might indicate that the model struggles to effectively detect and classify Infiltration attacks, especially with fewer features.**

**- The Web Attack category shows varying performance across different feature configurations, ranging from 3.774% to 87.342% accuracy. This suggests that the model has more difficulty in accurately detecting and classifying web attacks, especially with fewer features.**

**stacking using LSTM and CNN:**

**Accuracy:**

**LSTM\_CNN:**

**FAR:**

| **Attacks** | **15** | **24** | **48** | **ALL** |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| **Accuracy** | **99.482** | **99.144** | **99.425** | **99.417** |
| **FAR** | **0.518** | **0.86** | **0.575** | **0.58** |

**OVERALL PERFORMANCE COMPARISON:**

| **Model** | **Best Features** | **Accuracy** | **FAR** |
| --- | --- | --- | --- |
| **Naïve Bayes** | **24** | **0.76726** | **0.23274** |
| **KNN** | **ALL** | **0.99481** | **0.00519** |
| **Random Forest** | **24** | **0.99967** | **0.00033** |
| **XGB** | **48 and ALL** | **0.99981** | **0.00019** |
| **DCT** | **ALL** | **0.99953** | **0.00047** |
| **Bagging** | **ALL** | **0.99971** | **0.00029** |
| **Stacking using LSTM and CNN** |  |  |  |
| **Stacking** | **24** | **0.99979** | **0.00021** |
| **CNN** | **ALL** | **0.99546** | **0.0045** |
| **LSTM** | **24** | **0.99512** | **0.0049** |

**DCT FAR is 0.00071, if we apply boosting using the XGB algorithm then the FAR value reduces drastically. If we apply Bagging to the DCT it’s FAR value also reduces drastically. So, from the above results we can say that ensemble learning improved the performance.**

**Stacking also improved the performance of individual models.**

**Overall Best performance model is XGB with 48 and ALL features.**