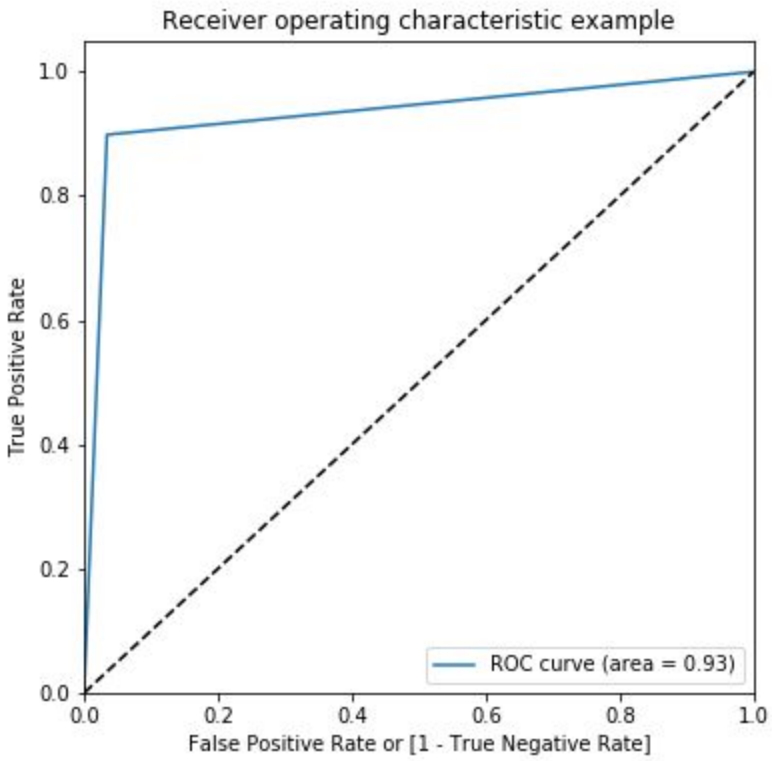
Lead Score Summary

According to the Problem Statement, there is an **educational company** which **sells online courses**, on various websites **to industry professionals**.
Any professional whoever goes through the courses are encountered with the form, if this **form is filled** with the individuals mailID and Phone Number then the **person is considered to be a "Lead" member**.
Once after these leads are sent to the marketing team of the company they make necessary conversation with the concerned person and try to convert them to a successful lead.
Overall **lead conversion rate of the company is very poor**. Hence the company wants to know **Hot Leads which will increase the lead conversion rate**.
Hence, the leads which are most likely to convert into paying leads have to be identified.
For doing so the following steps have been performed:

Given data,
Cleaning of the data was performed. **Removal of all the unnecessary columns** have been performed.
Columns which were having a **major** amount of **missing values** were **removed** as it would affect the models performance.
Then performing **Outlier detection** and removing outliers from the data. To this dataset finally **normalization** of 3 columns was performed and then **One-Hot-Encoding** on the categorical variable with more than 2 categories and **Replacing with 0/1** for categorical variables with 2 categories.
After the data preparation step, **Modelling** step was performed in this step there were multiple sub-steps which were performed.
Starting with **splitting of the data** into training set and testing set
Then Implementation of Logistic Regression on the entire data (131 columns) and finding the Correlation Matrix.
Removal of unnecessary variables after performing RFE Model and VIF model.
Rerunning the **Logistic Regression on the Filtered Variables (65 columns)**
Predicting the data with the "y_test" set and evaluated the results.
Drew **ROC curve and AUC Curve**.

ROC curve came out to be 0.93 while AUC came out to be 0.98. Which is a great model and can be used in production.



Conclusion:

Overall the model can be used by the company for predicting the leads and making those phone calls which are highly likely for the lead conversion and this will make the company work in the most efficient manner, reducing time spent on unnecessary calls, and achieving greater profit to the company.