**Statement of Purpose**

**My Introduction to machine-learning:**

Growing up I have always loved maths. I was that weird kid who always loved calculus, algebra and statistics. And because I excelled in the field of maths, I have always wanted to dive deep into the world of machine learning and datascience. I started with a course on Udemy and had watched various tutorials from YouTube channels like sentdex. Later, I realised that the real way of learning machine learning and datascience was to get hands-on on code rather than spending time hearing too many video lectures. So, I started out setting my environment with MiniConda and terminal. I started to get familiar with using tools like pandas (manipulating pandas series and dataframes – reading files and data from URLs, creating dataframes, viewing and handling dataframes of large datasets, dropping columns and entries , working with datasets with missing values and so on), NumPy (handling the shapes of inputs with reshape, dot product and transpose, creating, sorting and manipulating NumPy nd arrays, turning images into NumPy arrays, etc), matplotlib(I personally love visualising data and was deeply engrossed in this section for more than a week implementing scatter plots, bar plots, histograms, subplots, normal distributions, exporting my plots, customizing plots to look more appealing and comprehensive).

When it comes to the business end of machine learning projects, I love working with Scikit-learn and TensorFlow (deep learning). I enjoy the process of trying out different models and tuning hyperparameters to optimum levels. I have also tried to research a lot about the algorithms and their hyperparameters, trying to achieve good evaluation metrics. I usually follow a typical scikit-learn workflow-getting the data ready(clean, transform, reduce and train-test splitting), picking a model to suit the problem, fitting the model to the data, making predictions (with predict and predict\_proba), evaluating the model (with score method, roc curves, Area under the curve, classification reports with prediction, accuracy, recall and f1score, custom evaluation functions that make use of cross validation, etc), experimenting i.e. tuning hyperparameters manually or with randomizedsearchcv and gridsearchcv, and finally saving and loading the model using pickle or joblib. I am also interested in Exploratory Data Analysis (EDA) and feature importance that makes use of correlation matrices. I have been more focussed on Supervised learning with structured datasets.

**Project:**

Having honed my skills in all these domains, I had started working on projects. The first project I worked upon was a classification problem based on a heart-disease dataset to predict whether a patient has heart disease or not with a LogisticRegression model (have also tried other models) and ended up with a score of around 88% and a fairly good enough confusion matrix, cross-validation report(cross-val accuracy scores of 80-90%), roc curve, etc. The link to the project has been given below:

<https://github.com/senthil2000-dev/machinelearning/blob/master/heart_disease_classification.ipynb>

Next, I moved on to work on a regression problem, a bull-dozer sale price problem (time series data) -whose dataset was obtained from an old Kaggle competition. In this project I initially trained and optimised my model on a part of the dataset (since it was large) and then finally made a complete model on the full dataset. I made use of a RandomForestRegressor. I worked on it and was able to achieve a score that was good enough to fetch me a rank in the top 60 in the leader board (despite the competition having expired long ago). The link to the project has been appended below:

<https://github.com/senthil2000-dev/machinelearning/blob/master/bulldozer-price-regression.ipynb>

The dataset of the bulldozer prices has been retrieved from a Kaggle competition. Since it is a large dataset, I have placed a link below to it

<https://www.kaggle.com/c/bluebook-for-bulldozers/data>

Download all the data sources from this link and put it inside the data/bluebook-for-bulldozers to clone the GitHub repository.

Of late, I have learnt about deep learning and have worked on a dog breed prediction problem (whose dataset was also obtained from Kaggle) and was able to achieve a score which would get me a rank of around 30 on the Kaggle leader board. Since this was my first project based on unstructured data, I focussed more on basic concepts like turning data into tensors, turning data into batch format, applying transfer learning (MobileNetV2 from keras), turning predictions into text, using transformers, column, Onehotencoding, etc. This project is almost near completion and has few final optimisations to be done. I also have a brief knowledge of Big data analysis.

**My willingness to join the club and why?**

I was really fascinated by the trained model that the Delta members displayed during the induction-briefing process. So, I strongly believe that the Delta mentorship program would provide me an opportunity to dive deeper into the field of machine-learning, learn more about undiscovered topics like pytorch and pure deep learning algorithms (without transfer learning). Due to the nature of this mentorship program, I think the process of inductions will catalyse my learning process and moreover, I am also deeply interested in joining the club. I am also interested in learning new technologies everyday and adapt to various programming languages at the same time. So, I think I have found the right place to gather more knowledge in this domain and have a strong will to get inducted into the club. Not only do I wish to hone my skills and wish to become a machine learning and data science expert but also become a senior developer in this broad realm of programming and be able to deploy my machine learning projects on web and android apps (I always like to work on integrated projects).

**Concluding…**

My personal goals are to learn as much as I can about programming that would strengthen my awareness about trends in the field and how I can contribute to the overall mission and purpose of any organisation or club that I join. Thereby, wishing that induction would be a good learning process that would provide insights into exploring the usage of various ML models and tools, I hope for a positive reply of my application being accepted in this domain. Link to GitHub repository of my ML projects has been appended below:

<https://github.com/senthil2000-dev/machinelearning>