

```
In [6]: # Import our dependencies
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, OneHotEncoder, MinMaxScaler
import pandas as pd
import tensorflow as tf
import numpy as np

# Import our input dataset
df = pd.read_csv('../pitcher_salaries_cleaned.csv')
df.head()
```

Out[6]:

	Year	Full Name	Age	Salary	ERA	Hits	Earned Runs	Strike Outs	Home Runs	Wins	Losses	Outs Pitched	Batters Faced by Pitcher	Games Finished	Weight
0	1990	AbbottJim	23	185000	4.51	246	106	105	16	10	14	635	925	0	200
1	1990	AbbottPaul	23	100000	5.97	37	23	25	0	0	5	104	162	0	185
2	1990	AldredScott	22	100000	3.77	13	6	7	0	1	2	43	63	0	195
3	1990	AndersonAllan	26	300000	4.53	214	95	82	20	7	18	566	797	0	178
4	1990	AppierKevin	23	100000	2.76	179	57	127	13	12	8	557	784	1	180

Create Salary Brackets

```
In [10]: # Look at distribution of salaries (suppressing scientific notation)
df['Salary'].describe().apply(lambda x: format(x, 'f'))
```

Out[10]:

count	4937.000000
mean	3011304.443387
std	4265619.190449
min	100000.000000
25%	327000.000000
50%	980000.000000
75%	4000000.000000
max	33000000.000000

Name: Salary, dtype: object

```
In [24]: # create salary brackets and labels
bins = [0, 499999, 4999999, 9999999, 34999999]
labels = ['low', 'mid', 'high', 'top']
```

```
In [32]: # apply salary brackets
df['Salary Bin'] = pd.cut(df['Salary'], bins=bins, labels=labels)
df
```

Out[32]:

	Year	Full Name	Age	Salary	ERA	Hits	Earned Runs	Strike Outs	Home Runs	Wins	Losses	Outs Pitched	Batters Faced by Pitcher	Games Finished
0	1990	AbbottJim	23	185000	4.51	246	106	105	16	10	14	635	925	0
1	1990	AbbottPaul	23	100000	5.97	37	23	25	0	0	5	104	162	0
2	1990	AldredScott	22	100000	3.77	13	6	7	0	1	2	43	63	0
3	1990	AndersonAllan	26	300000	4.53	214	95	82	20	7	18	566	797	0
4	1990	AppierKevin	23	100000	2.76	179	57	127	13	12	8	557	784	1

	Year	Full Name	Age	Salary	ERA	Hits	Earned Runs	Strike Outs	Home Runs	Wins	Losses	Outs Pitched	Batters Faced by Pitcher	Games Finished
...
4932	2016	WorleyVance	29	2600000	3.53	84	34	56	11	2	2	260	365	13
4933	2016	WrightMike	26	510500	5.79	81	48	50	12	3	4	224	328	5
4934	2016	WrightSteven	32	514500	3.33	138	58	127	12	13	6	470	656	0
4935	2016	YoungChris	37	4250000	6.19	104	61	94	28	3	9	266	406	7
4936	2016	ZimmermannJordan	30	18000000	4.87	118	57	66	14	9	7	316	450	1

4937 rows × 15 columns



Encode Salary Bins column

In [39]:

```
# encode object features
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
encoded_df = df.copy()
df['Salary Bin'] = le.fit_transform(df['Salary'])

df.head()
```

Out[39]:

	ERA	Hits	Earned Runs	Strike Outs	Home Runs	Wins	Losses	Outs Pitched	Batters Faced by Pitcher	Games Finished	Weight	Height	Games Started	Salary Bin
0	4.51	246	106	105	16	10	14	635	925	0	200	75	33	1
1	5.97	37	23	25	0	0	5	104	162	0	185	75	7	1
2	3.77	13	6	7	0	1	2	43	63	0	195	76	3	1
3	4.53	214	95	82	20	7	18	566	797	0	178	71	31	1
4	2.76	179	57	127	13	12	8	557	784	1	180	74	24	1

In [33]:

```
# drop unnecessary columns
df = df.drop(["Full Name", "Team", "League", "Age", "Year", "Salary"], 1)
df.head()
```

C:\Users\alyss\anaconda3\envs\mlenv\lib\site-packages\ipykernel_launcher.py:1: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only
 """Entry point for launching an IPython kernel.

Out[33]:

	ERA	Hits	Earned Runs	Strike Outs	Home Runs	Wins	Losses	Outs Pitched	Batters Faced by Pitcher	Games Finished	Weight	Height	Games Started	Salary Bin
0	4.51	246	106	105	16	10	14	635	925	0	200	75	33	low
1	5.97	37	23	25	0	0	5	104	162	0	185	75	7	low
2	3.77	13	6	7	0	1	2	43	63	0	195	76	3	low
3	4.53	214	95	82	20	7	18	566	797	0	178	71	31	low
4	2.76	179	57	127	13	12	8	557	784	1	180	74	24	low

Split Features/Target & Training/Testing Sets

Split into features and target

- **y variable:** Our target variable, Salary
- **X variable:** Our features; just drop Salary and Full Name

In [40]:

```
# Split our preprocessed data into our features and target arrays
y = df["Salary Bin"].values
X = df.drop(["Salary Bin"],1).values

# Split the preprocessed data into a training and testing dataset
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1)
```

C:\Users\alyss\anaconda3\envs\mlenv\lib\site-packages\ipykernel_launcher.py:3: FutureWarning: In a future version of pandas all arguments of DataFrame.drop except for the argument 'labels' will be keyword-only
This is separate from the ipykernel package so we can avoid doing imports until

Build and Instantiate StandardScaler object, then standardize numerical features

In [41]:

```
# Create a StandardScaler instance
scaler = StandardScaler()

# Fit the StandardScaler
X_scaler = scaler.fit(X_train)

# Scale the data
X_train_scaled = X_scaler.transform(X_train)
X_test_scaled = X_scaler.transform(X_test)
```

Build Neural Net Framework

In [42]:

```
# Define the model - deep neural net
number_input_features = len(X_train[0])
hidden_nodes_layer1 = 50
hidden_nodes_layer2 = 40
hidden_nodes_layer3 = 30

nn = tf.keras.models.Sequential()

# First hidden layer
nn.add(
    tf.keras.layers.Dense(units=hidden_nodes_layer1, input_dim=number_input_features, activation="tanh")
)

# Second hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer2, activation="relu"))

# Third hidden layer
nn.add(tf.keras.layers.Dense(units=hidden_nodes_layer3, activation="relu"))

# Output layer
nn.add(tf.keras.layers.Dense(units=1, activation="sigmoid"))

# Check the structure of the model
nn.summary()
```

Model: "sequential_1"

Layer (type)	Output Shape	Param #
=====		

dense_4 (Dense)	(None, 50)	700
dense_5 (Dense)	(None, 40)	2040
dense_6 (Dense)	(None, 30)	1230
dense_7 (Dense)	(None, 1)	31

=====

Total params: 4,001
Trainable params: 4,001
Non-trainable params: 0

Compile the Model

In [47]:

```
# Compile the model
nn.compile(loss="CategoricalCrossentropy", optimizer="adam", metrics=["accuracy"])
```

Train the model

In [48]:

```
# Train the model
fit_model = nn.fit(X_train,y_train,epochs=200)
```

```
Epoch 1/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3712
Epoch 2/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 3/200
116/116 [=====] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 4/200
116/116 [=====] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 5/200
116/116 [=====] - 0s 670us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 6/200
116/116 [=====] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 7/200
116/116 [=====] - 0s 661us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 8/200
116/116 [=====] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 9/200
116/116 [=====] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 10/200
116/116 [=====] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 11/200
116/116 [=====] - 0s 670us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 12/200
116/116 [=====] - 0s 643us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 13/200
116/116 [=====] - 0s 643us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 14/200
116/116 [=====] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 15/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 16/200
116/116 [=====] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 17/200
116/116 [=====] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 18/200
116/116 [=====] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 19/200
116/116 [=====] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 20/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 21/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 22/200
116/116 [=====] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 23/200
```

116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 24/200									
116/116 [=====]	-	0s	635us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 25/200									
116/116 [=====]	-	0s	652us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 26/200									
116/116 [=====]	-	0s	687us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 27/200									
116/116 [=====]	-	0s	670us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 28/200									
116/116 [=====]	-	0s	635us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 29/200									
116/116 [=====]	-	0s	670us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 30/200									
116/116 [=====]	-	0s	635us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 31/200									
116/116 [=====]	-	0s	678us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 32/200									
116/116 [=====]	-	0s	635us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 33/200									
116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 34/200									
116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 35/200									
116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 36/200									
116/116 [=====]	-	0s	635us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 37/200									
116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 38/200									
116/116 [=====]	-	0s	635us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 39/200									
116/116 [=====]	-	0s	652us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 40/200									
116/116 [=====]	-	0s	678us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 41/200									
116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 42/200									
116/116 [=====]	-	0s	644us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 43/200									
116/116 [=====]	-	0s	617us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 44/200									
116/116 [=====]	-	0s	643us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 45/200									
116/116 [=====]	-	0s	643us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 46/200									
116/116 [=====]	-	0s	617us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 47/200									
116/116 [=====]	-	0s	635us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 48/200									
116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 49/200									
116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 50/200									
116/116 [=====]	-	0s	626us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 51/200									
116/116 [=====]	-	0s	617us/step	-	loss:	0.0000e+00	-	accuracy:	0.3787
Epoch 52/200				</					

[illegible]

[illegible]

[illegible]

116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 176/200
116/116 [=====] - 0s 626us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 177/200
116/116 [=====] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 178/200
116/116 [=====] - 0s 644us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 179/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 180/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 181/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 182/200
116/116 [=====] - 0s 643us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 183/200
116/116 [=====] - 0s 643us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 184/200
116/116 [=====] - 0s 661us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 185/200
116/116 [=====] - 0s 670us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 186/200
116/116 [=====] - 0s 670us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 187/200
116/116 [=====] - 0s 687us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 188/200
116/116 [=====] - 0s 643us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 189/200
116/116 [=====] - 0s 617us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 190/200
116/116 [=====] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 191/200
116/116 [=====] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 192/200
116/116 [=====] - 0s 644us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 193/200
116/116 [=====] - 0s 644us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 194/200
116/116 [=====] - 0s 774us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 195/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 196/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 197/200
116/116 [=====] - 0s 635us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 198/200
116/116 [=====] - 0s 652us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 199/200
116/116 [=====] - 0s 670us/step - loss: 0.0000e+00 - accuracy: 0.3787
Epoch 200/200
116/116 [=====] - 0s 704us/step - loss: 0.0000e+00 - accuracy: 0.3787

In []: