Motivation

In this work was compared four different models for names classification by gender.

Data

Data represents by table with two columns (name and gender) and 83 288 rows,

Example:

Name	Gender
Terrone	M
Annaley	F

RNN models

Was trained few RNN models with different hyperparameters, the following results were obtained:

1. LSTM layer (size 16), 1 Dense Layer (size 16), optimizer adam with default parameter's. Overfiring starts after 80 epoch's, Training Accuracy: 0.8187 and Testing Accuracy: 0.8142

2. LSTM layer (size 16), Dropout(0.2), 1 Dense Layer (size 16), optimizer adam with default parameter's, 100 epoch's. Training Accuracy: 0.8238 and Testing Accuracy: 0.8181

3. GRU layer (size 18), Dropout(0.15), 1 Dense Layer (size 18), optimizer adam with default parameter's, 100 epoch's. Training Accuracy: 0.8155 and Testing Accuracy: 0.8116

Training of all three models was stopped before they start overfitting, usually they start overfit after 80^{th} epoch. Every model has about 2 400 parameters and best one is model with 1 LSTM layer (size 16), 1 Dense Layer (size 16)

Not RNN models

For comparison RNN model with other types of models was trained and the following results were obtained:

Logistic Regression - Testing Accuracy: 0.6287

1 Dense layer (size 15), optimizer adam with default parameter's, 50 epoch's. Training Accuracy: 0.8096 and Testing Accuracy: 0.8043

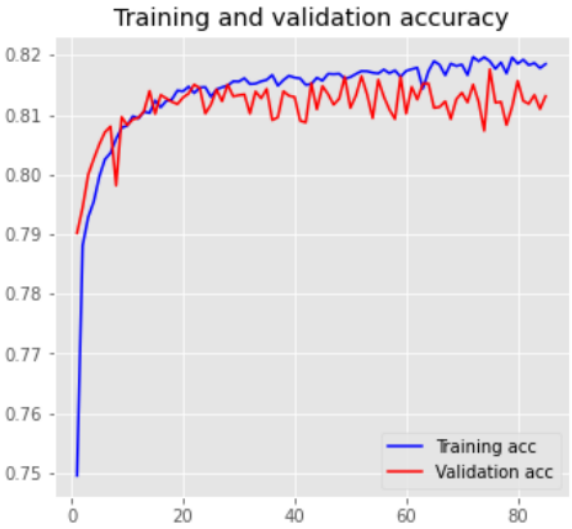
1 Global Max Pooling, 2 Dense layer (size 35), optimizer adam with default parameter's, 50 epoch's. Training Accuracy: 0.7053 and Testing Accuracy: 0.7017

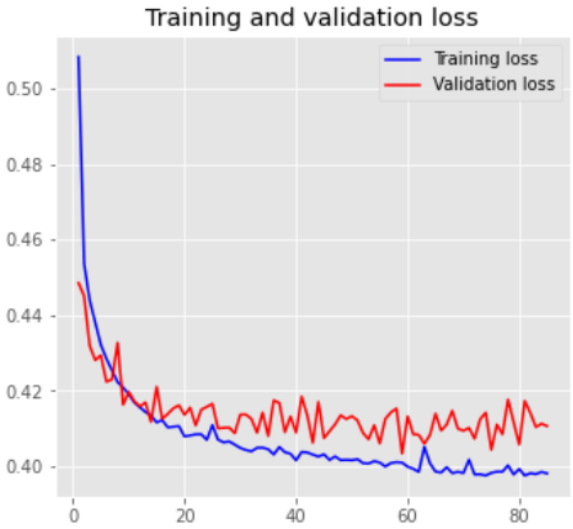
All three of these models have the same number of paramiters as RNN models – 2450 +- 100, overfitting starts after 50 epochs. Heres the best model is fully connected NN.

Graphs and conclusion

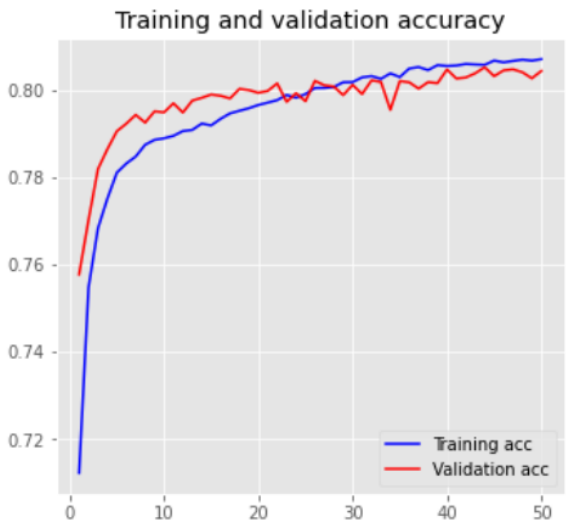
In this paragraph presented graphs for 3 models:

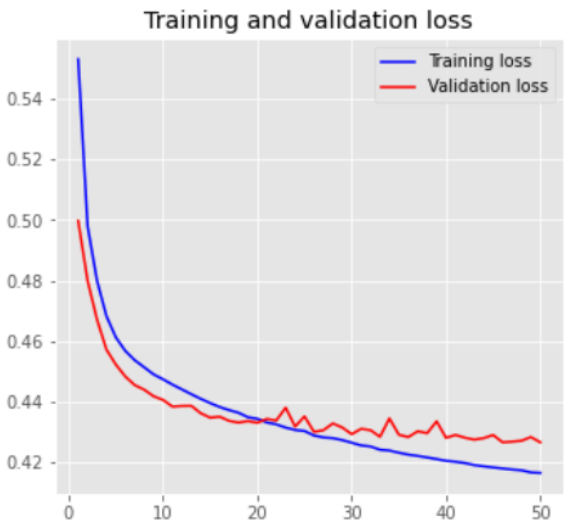
model with 1 LSTM layer (size 16), 1 Dense Layer (size 16)



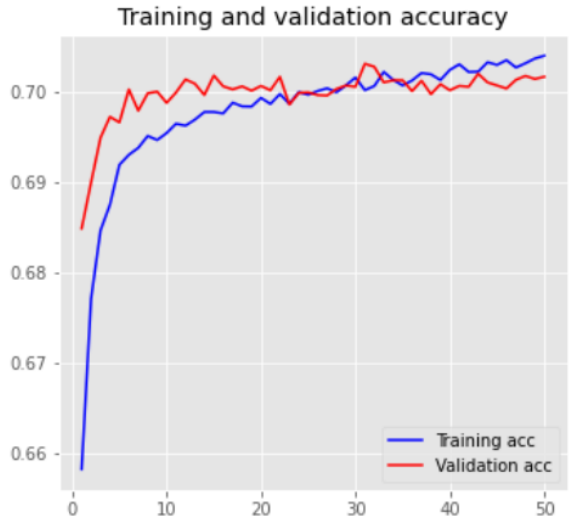


1 Dense layer (size 15)





Global Max Pooling, 2 Dense layer (size 35)



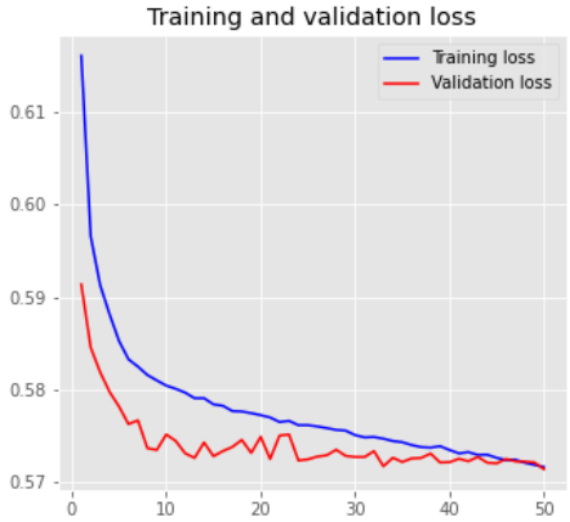
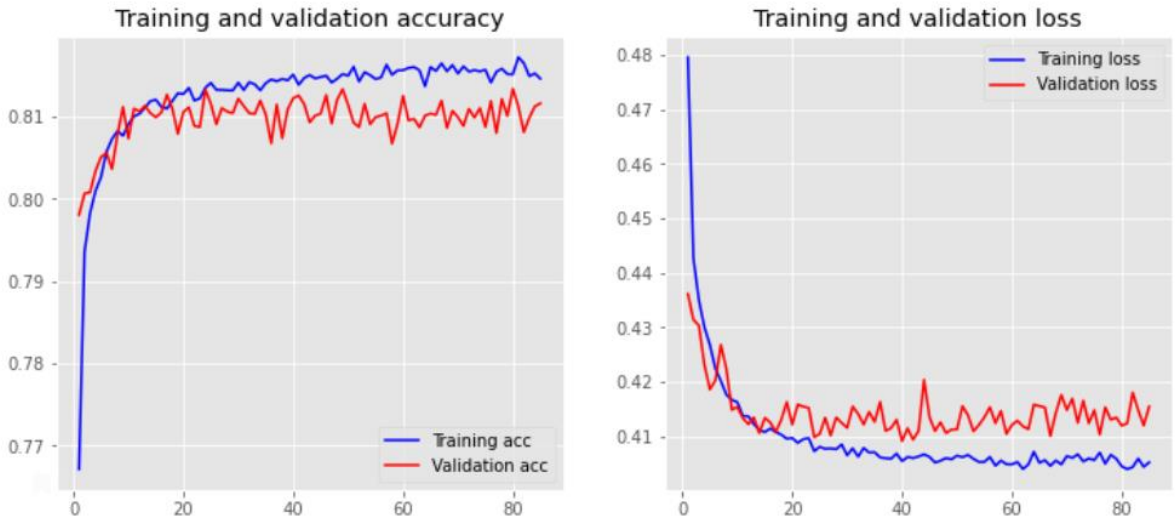


Table for comparison these three models:

Model	Train accuracy	Test accuracy
LSTM	0.8206	0.8131
Dense	0.8096	0.8043
Max pooling and Dense	0.7053	0.7017

From this table we can see that LSTM based model is better in terms of accuracy. Nevertheless, if we reduce number of parameters LSTM model will compute faster and still have same accuracy as fully connected model.

Here is model with almost twice less parameters, 1 GRU layer (15 size). Training Accuracy: 0.8167 and Testing Accuracy: 0.8116



At the end we found what for names classification task is better to use RNN model especially with GRU layers, this model can have not much parameters and rich 0.81 accuracy.