FAKE NEWS CLASSIFIER

BENJAMIN ANG
ALEXANDER PINO
DANE ROSARIO
JEREMY TAN

AGENDA OF THE PRESENTATION

01 INTRODUCTION

02 DATASET AND METHODOLOGY

03 RESULTS AND ANALYSIS

04 CONCLUSION, COMPARISON, AND RECCOMENDATIONS

Definitions

- News that is verifiably false
- Misleading information
- Deliberately/unintentionally misinforms or deceives readers
- Manipulates perception

FAKE NEWS IN THE PHILIPPINES

01 It's become much harder to filter content especially since fake news and misinformation have become prevalent during the pandemic.

02 Viral spread of false information has serious implications on the behaviours, attitudes and beliefs of the public, and ultimately can seriously endanger the democratic processes.

03 Fake news tends to disproven pseudoscience – e.g. anti-vax movement

WHAT IS FAKE NEWS?

"Fake news" are those news stories that are false: the story itself is fabricated, with no verifiable facts, sources or quotes.

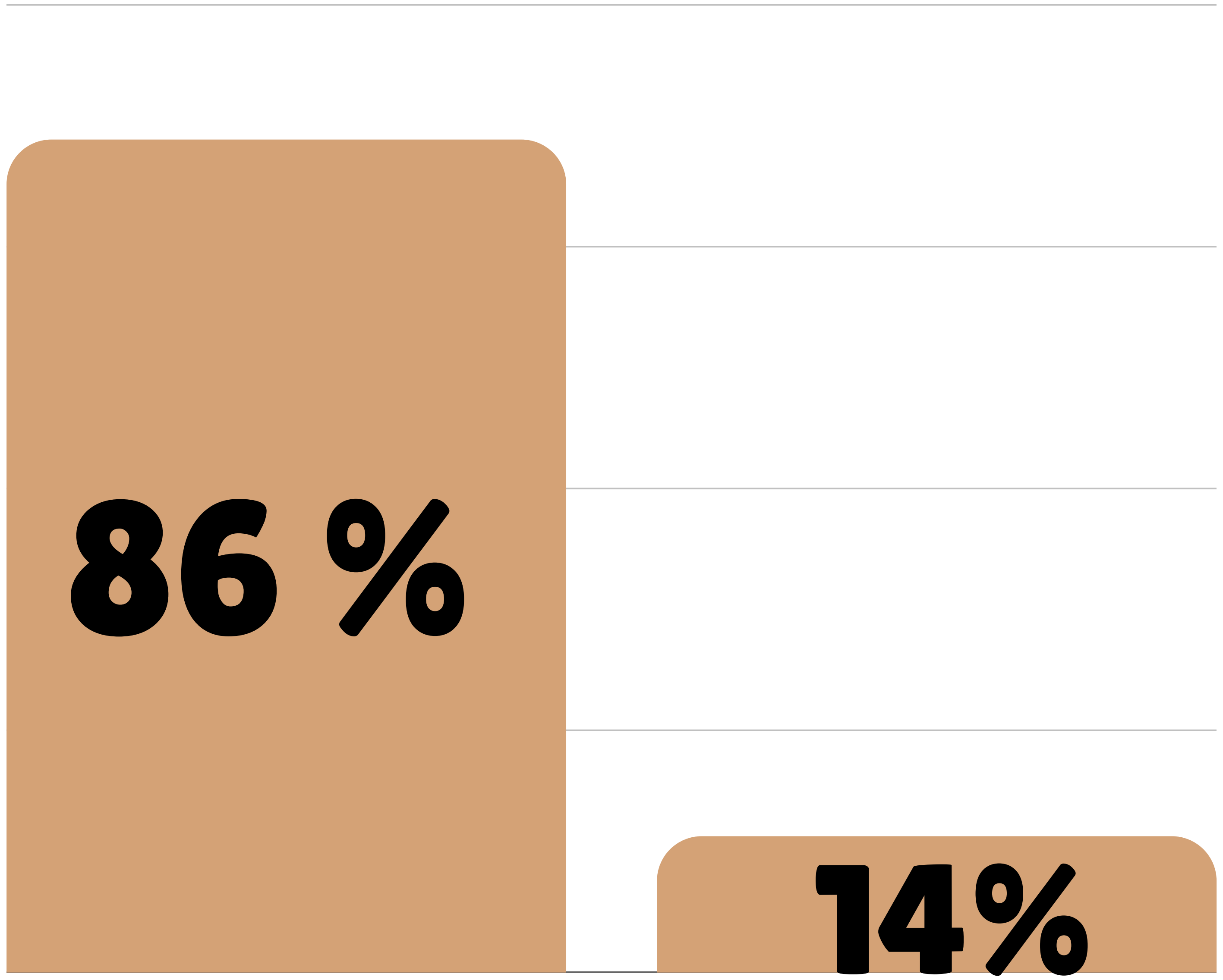
BY THE NUMBERS

1200 respondents were surveyed by Pulse Asia on September 2022 and 86% said that fake news is indeed a problem in the Philippines

9 IN 10 Filipinos view fake news as a major problem online. (Pulse Asia)

68 % claimed that they have interacted with fake news mostly on the internet

58 % pointed out that social media influencers or bloggers are the top sources of fake news



OUR DATASET

LOW-RESOURCE FAKE NEWS DETECTION CORPORA IN FILIPINO

The first of its kind. Contains 3,206 expertly-labeled news samples, half of which are real and half of which are fake (Cruz et al., 2020)

"FAKE NEWS DETECTION IN FILIPINO"

SAMPLE DATA

LABEL	ARTICLE	SOURCE OF NEWS
0 (Real)	Ayon sa TheWrap.com, naghain ng kaso si Krupa, 35, noong Huwebes dahil nakaranas umano siya ng emotional distress bunga ng mga malisyosong pahayag ni......	Mainstream news websites e.g. Pilipino Star Ngayon, Balita
1 (Fake)	Nagbigay na rin ng opinyon ang mga mix martial artists at organizers sa bansa kaugnay sa kumalat na video ng batang atenista na nambubugbog ng kapwa	"Fake news websites" tagged by NUJP, Verafiles

FEATURES OF FILIPINO LANGUAGE

- **Low-resource**, lacking large and expertly-produced datasets for training text models,
- **Morphologically-rich**, with diversity in the internal structure of words:
  - Prefixes (e.g. *Mag*-hintay)
  - Infixes (e.g. P-*in*-aasa)
  - Suffixes (e.g. Balik-*an*)
  - Compound words (e.g. *kathang-isip*)
- **Flexible sentence structure**
  - Verb before subject
  - Subject before verb

(Cruz et al., 2020)

METHODOLOGY

OBJECTIVE

Predict if a given news article is **Real** or **Fake**

PRE-PROCESS

Byte Pair Encoding tokenizer
One-hot Sequence Encoding
Vocabulary (20k), sequence length (200 subwords)

FIT

Dense NN (2 layers)
Long Short-term Memory NN (bidirectional, dropout)
Convolutional NN (single 1-D convolutional layer, slider)

EVALUATE

Accuracy, ROC graph
Final evaluation on Testing split

LOCALIZATION OF FAKE NEWS DETECTION VIA MULTITASK TRANSFER LEARNING (CRUZ 2020)

PRE-PROCESS

Byte Pair Encoding tokenizer

One-hot Sequence Encoding

Vocabulary (20k), sequence length (200 subwords)

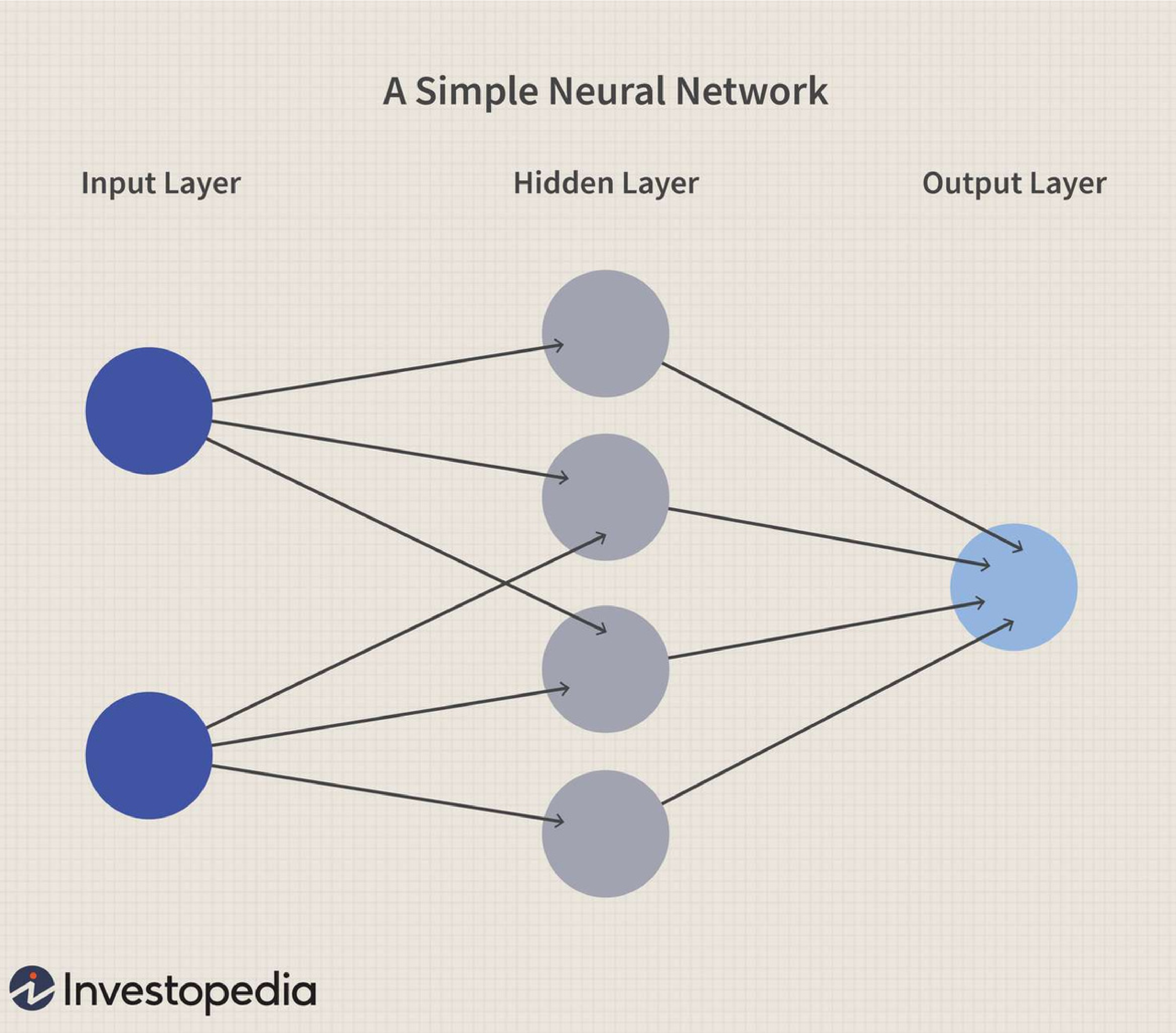
Split	Label	Documents	Tokens	Unique Tokens	OOV Tokens
Train	-	2,244	468,056	41,570	
Test	-	962	200,472	24,978	8,807 (17.48%)
Train + Test	Real	1,603	447,401	35,959	
	Fake	1,603	221,127	27,371	14,418 (28.62%)
Train	Real	1,121	312,047	29,619	
	Fake	1,123	156,009	22,385	11,951 (28.75%)

RESULTS

HERE ARE THE RESULTS FROM THE THREE NEURAL NETWORK MODELS

THIS WILL COVER:

- relevant parameters used
- recipe and pre-processing steps
- metrics and model evaluation



DENSE NEURAL NETWORKS

WHAT IS IT?

- Layers of nodes take input from previous layers adjusted by weight
- Most basic neural network model for classification tasks
- Converts sequences of tokens of fixed length into a numerical array of indices
- The output layer has an activation function to output the predicted probability

MODEL PARAMETERS AND RECIPE PRE PROCESSING STEPS

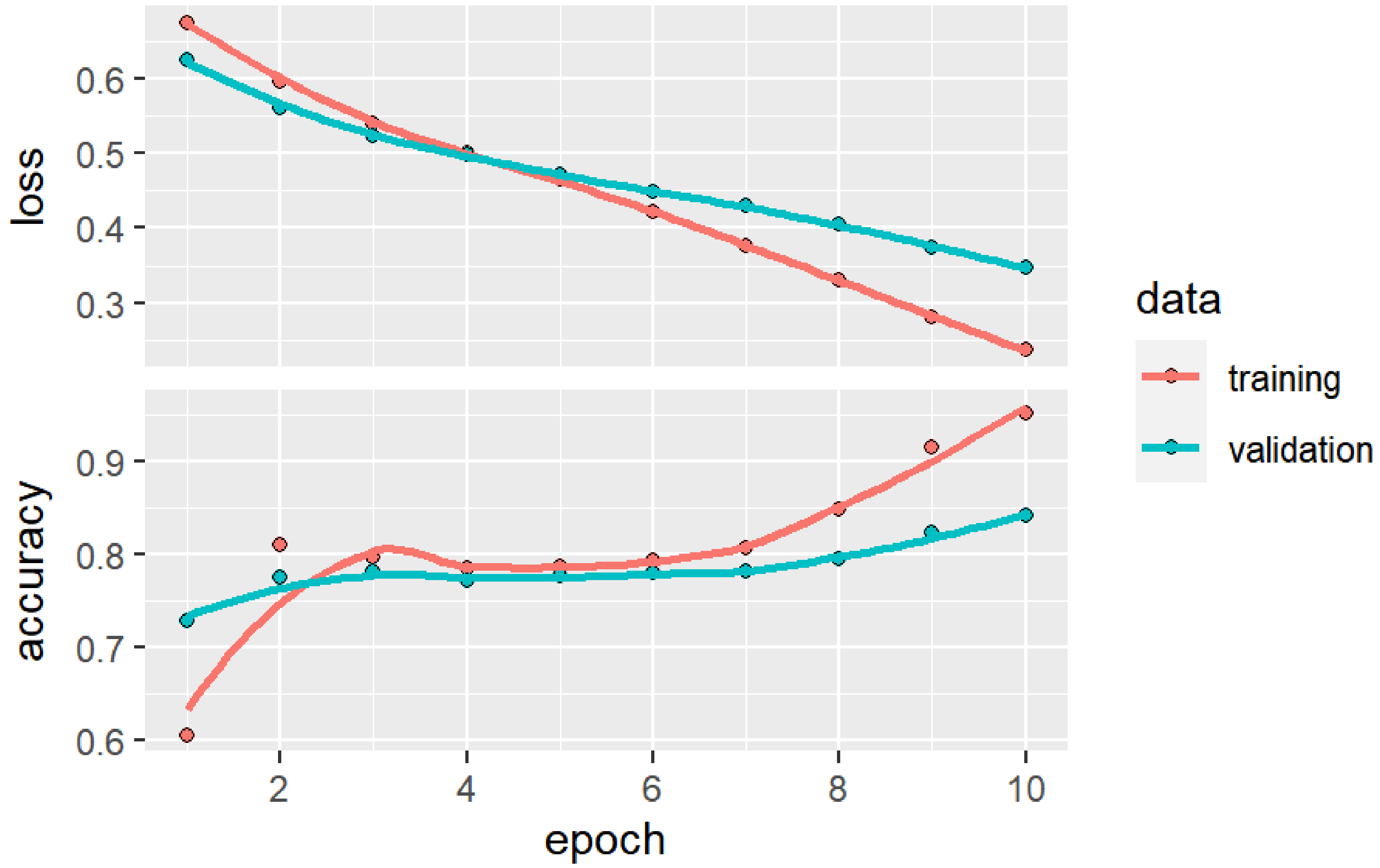
````r
max_subwords <- 2e4
max_length <- 200
news_rec <- recipe(~article, data = news_train) %>%
  step_mutate(article = tolower(article)) %>%
  step_tokenize(article,
                engine = "tokenizers.bpe",
                training_options = list(vocab_size =
max_subwords)) %>%
  step_sequence_onehot(article, sequence_length =
max_length)
news_rec
```
````
````

Dense Neural Network Architecture

```
Process time: 0.13 sec elapsed
Model: "sequential_3"

_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_3 (Embedding     (None, 200, 12)           240012
 )
 flatten_3 (Flatten)        (None, 2400)              0
 dense_7 (Dense)            (None, 32)                76832
 dense_6 (Dense)            (None, 1)                 33
=================================================================
Total params: 316,877
Trainable params: 316,877
Non-trainable params: 0
```

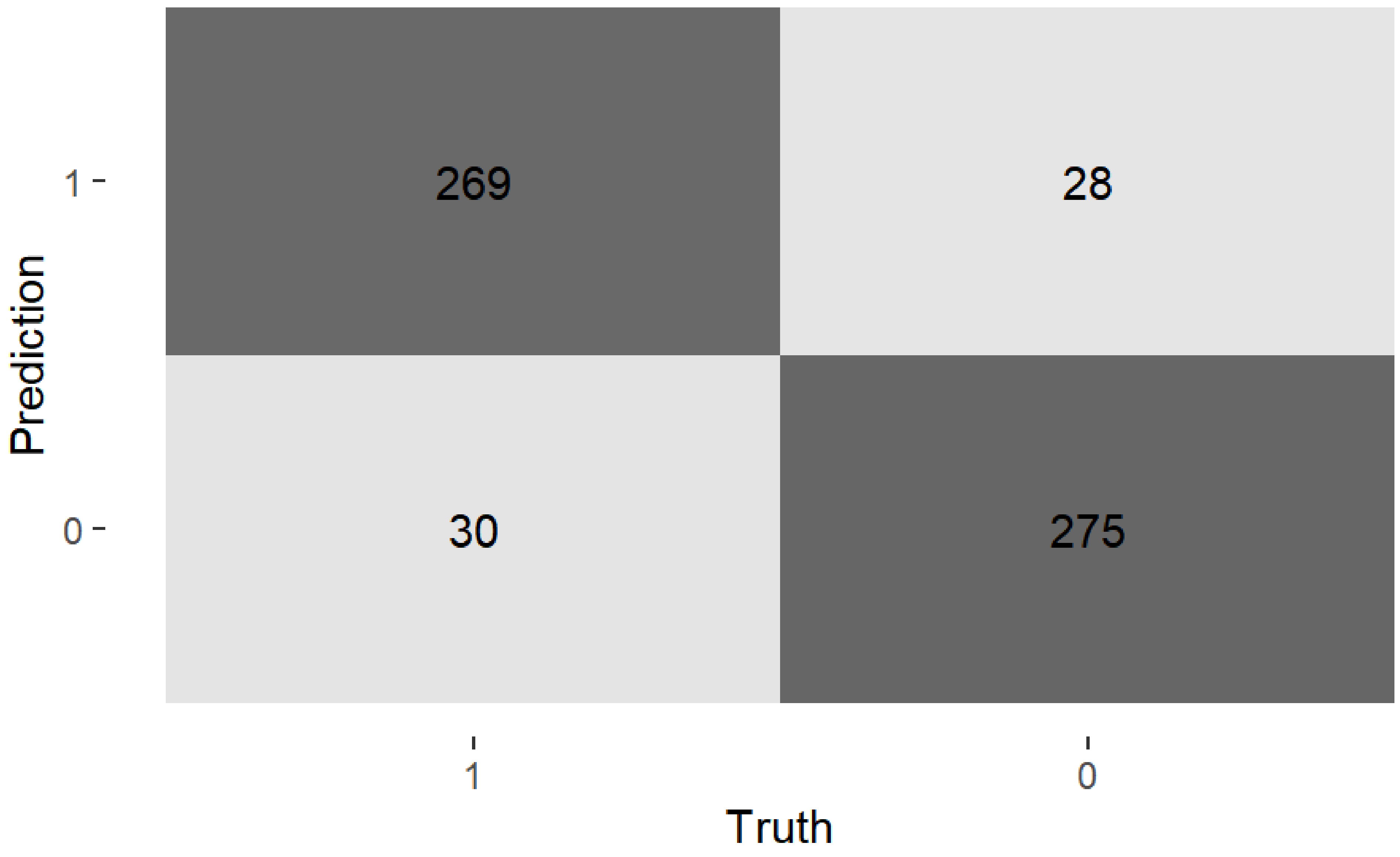
ACCURACY AND LOSS METRICS OVER 10 EPOCHS




```
Final epoch (plot to see history):
        loss: 0.2366
    accuracy: 0.9517
    val_loss: 0.3473
val_accuracy: 0.8405
```

DENSE NEURAL NETWORKS

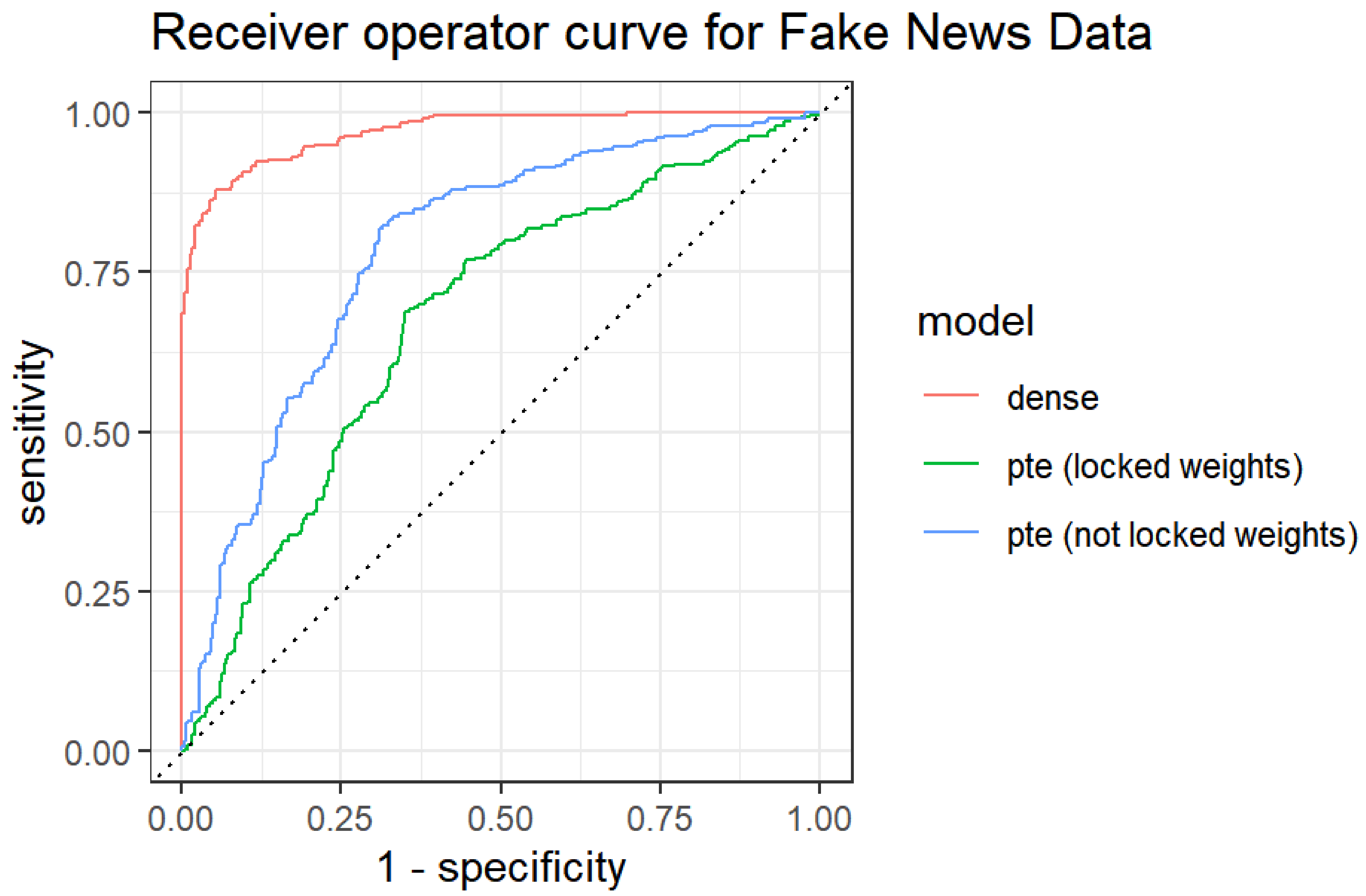
CONFUSION MATRIX ON OUR INITIAL VALILDATION SET




```
Final epoch (plot to see history):
        loss: 0.2366
    accuracy: 0.9517
    val_loss: 0.3473
val_accuracy: 0.8405
```

BINARY CROSSENTROPY LOSS FUNCTION

ROC CURVE



Receiver operator curve for Fake News Data
1.00
0.75
0.50
0.25
0.00
sensitivity
0.00
0.25
0.50
0.75
1.00
1 - specificity
model
dense
pte (locked weights)
pte (not locked weights)

REGULAR DNN PERFORMED THE BEST (SELF LEARNED EMBEDDINGS)

A tibble: 6 × 4

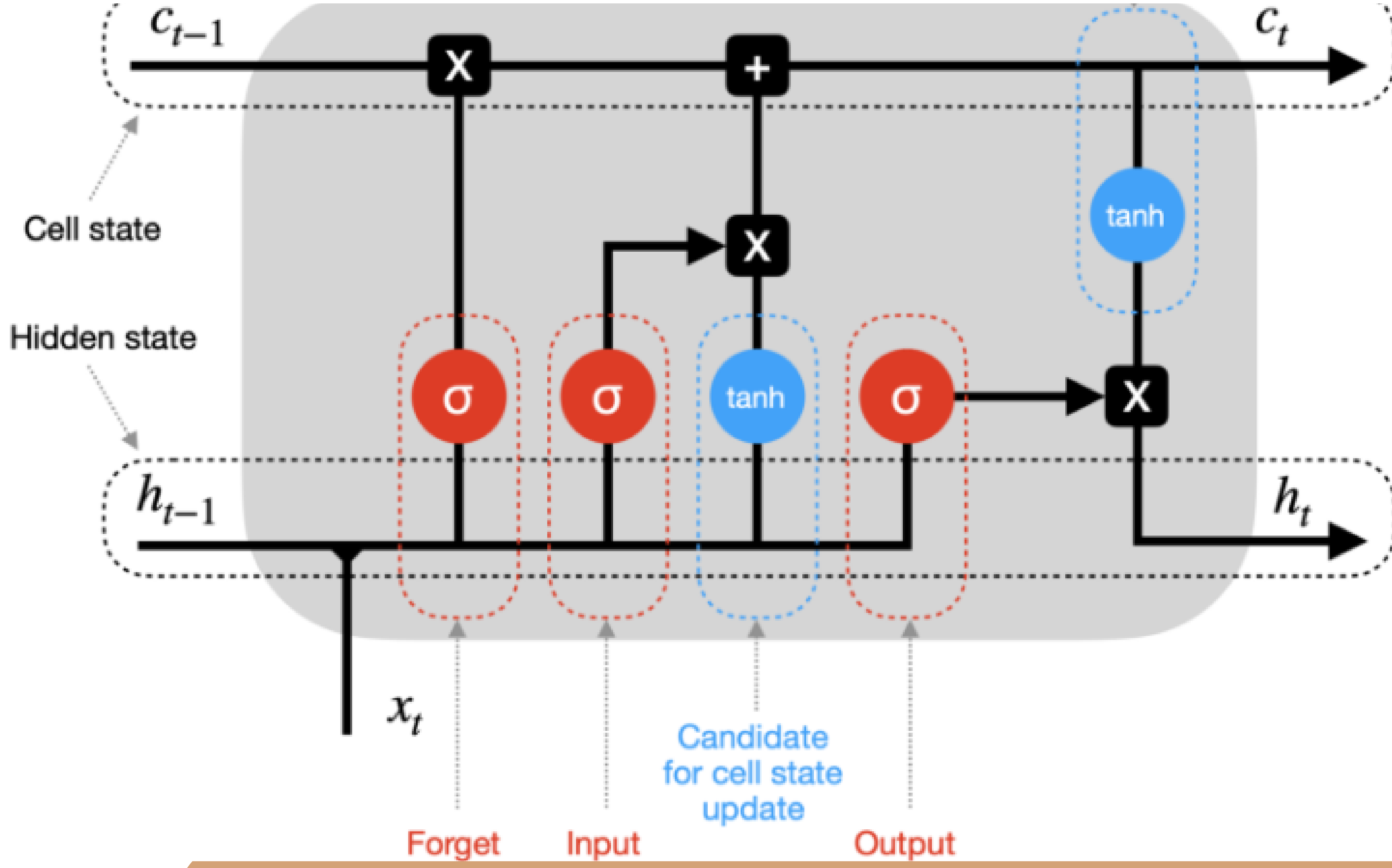
model <chr>	.metric <chr>	.estimator <chr>	.estimate <dbl>
dense	accuracy	binary	0.9036545
pte (locked weights)	accuracy	binary	0.6362126
pte (not locked weights)	accuracy	binary	0.7392027
dense	kap	binary	0.8072920
pte (locked weights)	kap	binary	0.2721521
pte (not locked weights)	kap	binary	0.4786930

EVALUATION ON THE TESTING DATA

.metric <chr>	.estimator <chr>	.estimate <dbl>
accuracy	binary	0.9002494
kap	binary	0.8002640
mn_log_loss	binary	0.2376600
roc_auc	binary	0.9652439

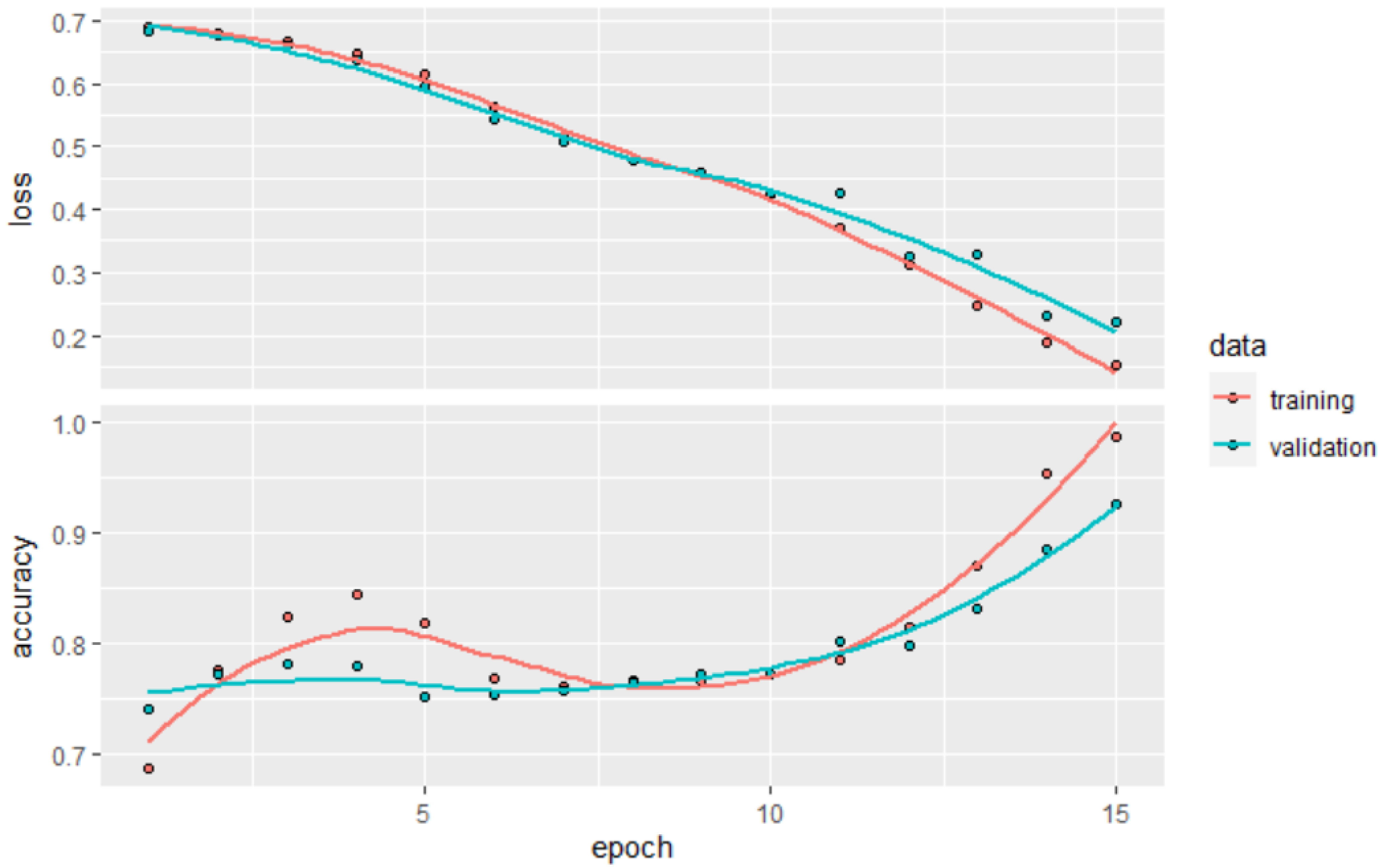
LONG SHORT TERM MEMORY (LSTM)

- Unlike conventional neural networks, LSTM can retain information from earlier parts of the sequence
- It does this by allowing each node to keep track of the cell state (memory cells)
- Capable of bi-directionality - reading sequences of texts backwards
- Dropout - removes certain tokens to reduce overfitting to training data



ACCURACY AND LOSS METRICS OVER 15 EPOCHS

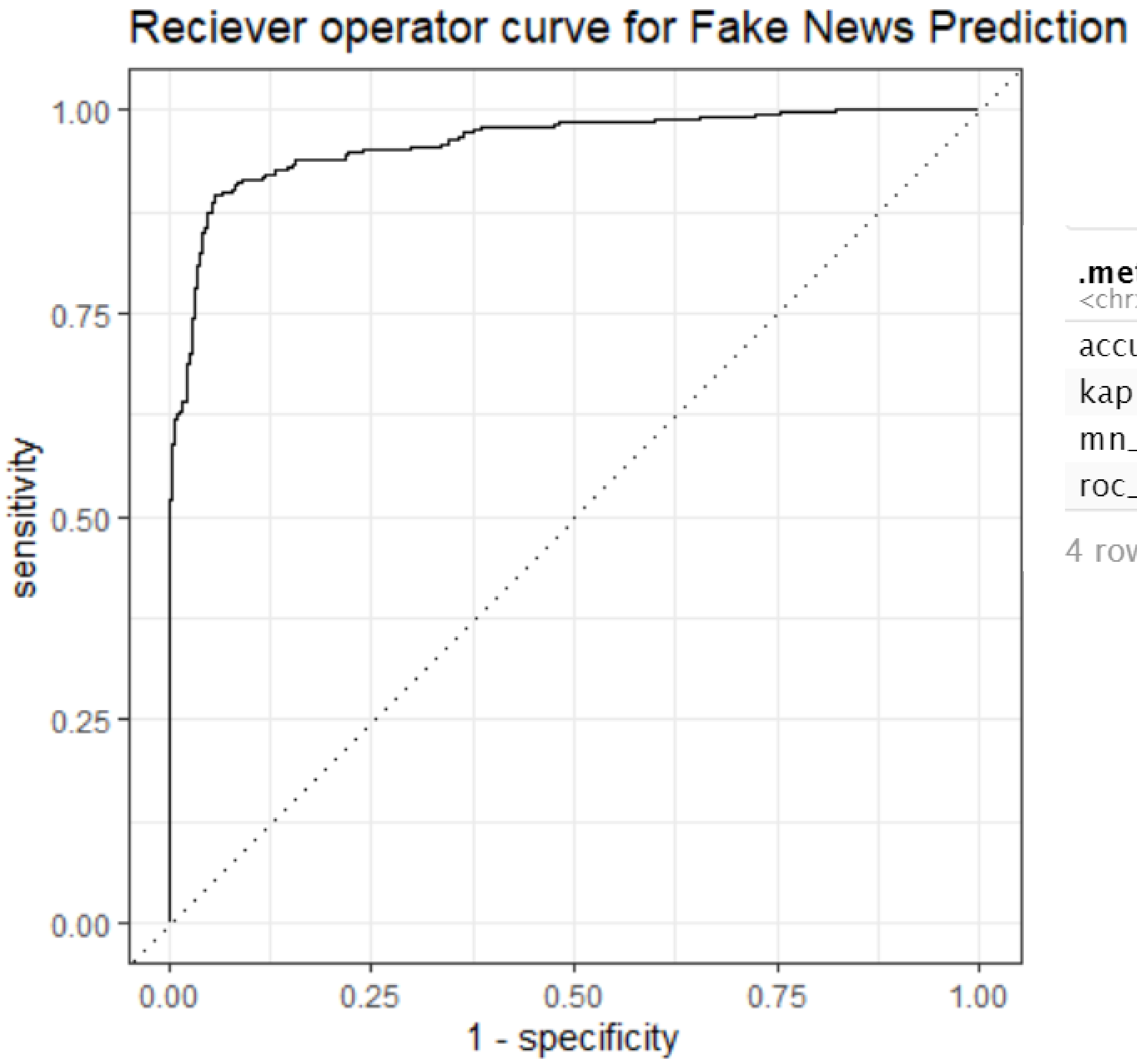
LSTM



CONFUSION MATRIX ON OUR INITIAL VALILDATION SET

LSTM

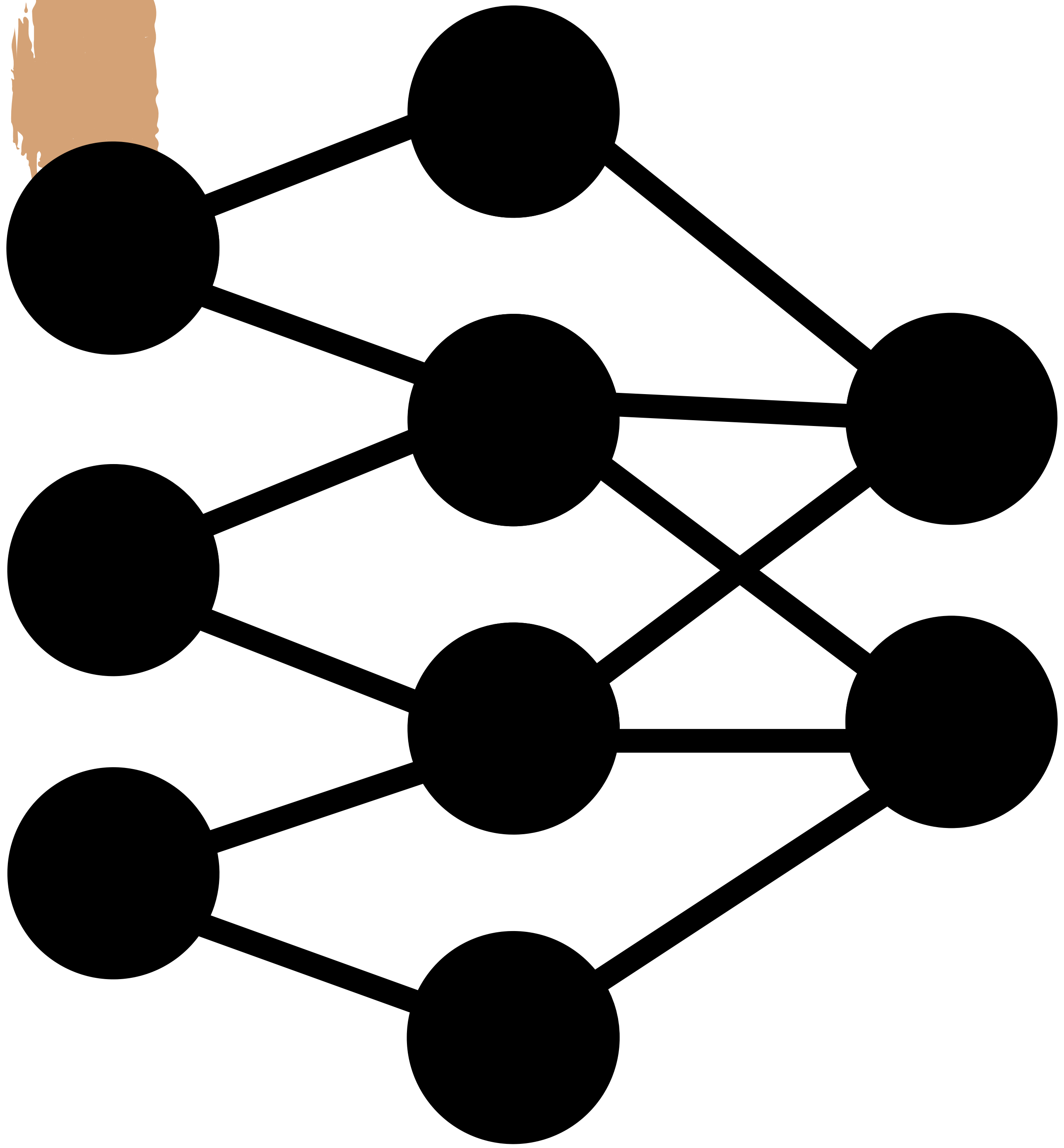
ROC CURVE



.metric <chr>	.estimator <chr>	.estimate <dbl>
accuracy	binary	0.9127182
kap	binary	0.8254310
mn_log_loss	binary	0.2788496
roc_auc	binary	0.9593595

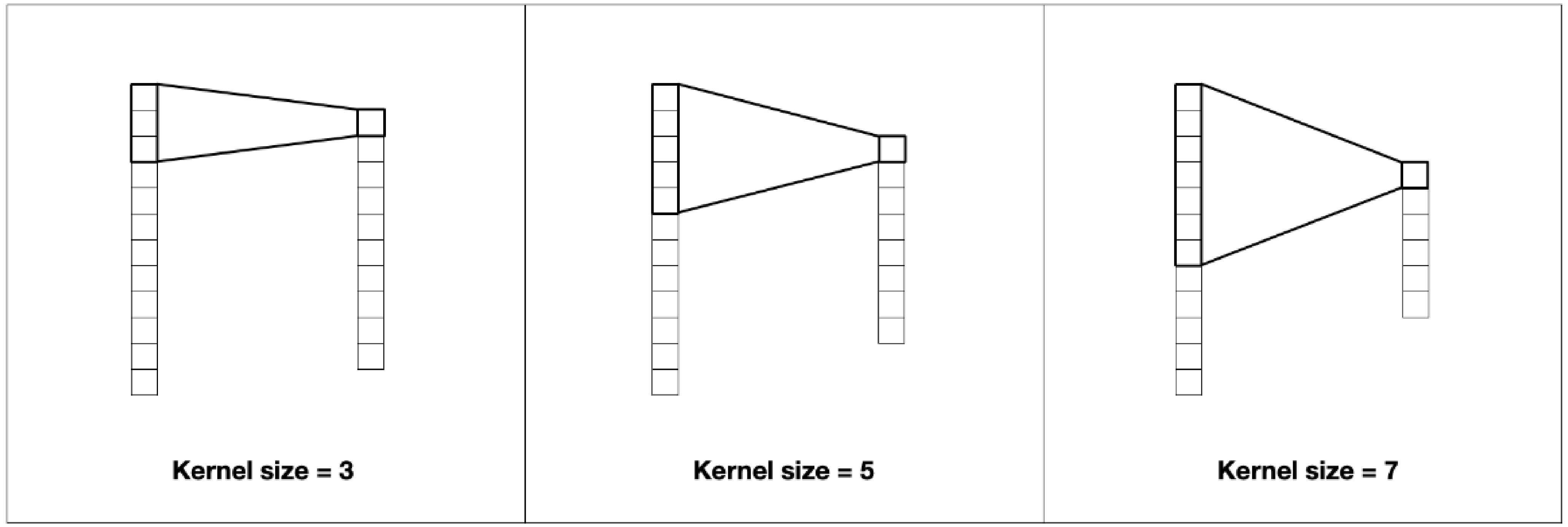
4 rows

CONVOLUTIONAL NEURAL NETWORKS



Specializes in analyzing the **spatial** location of words

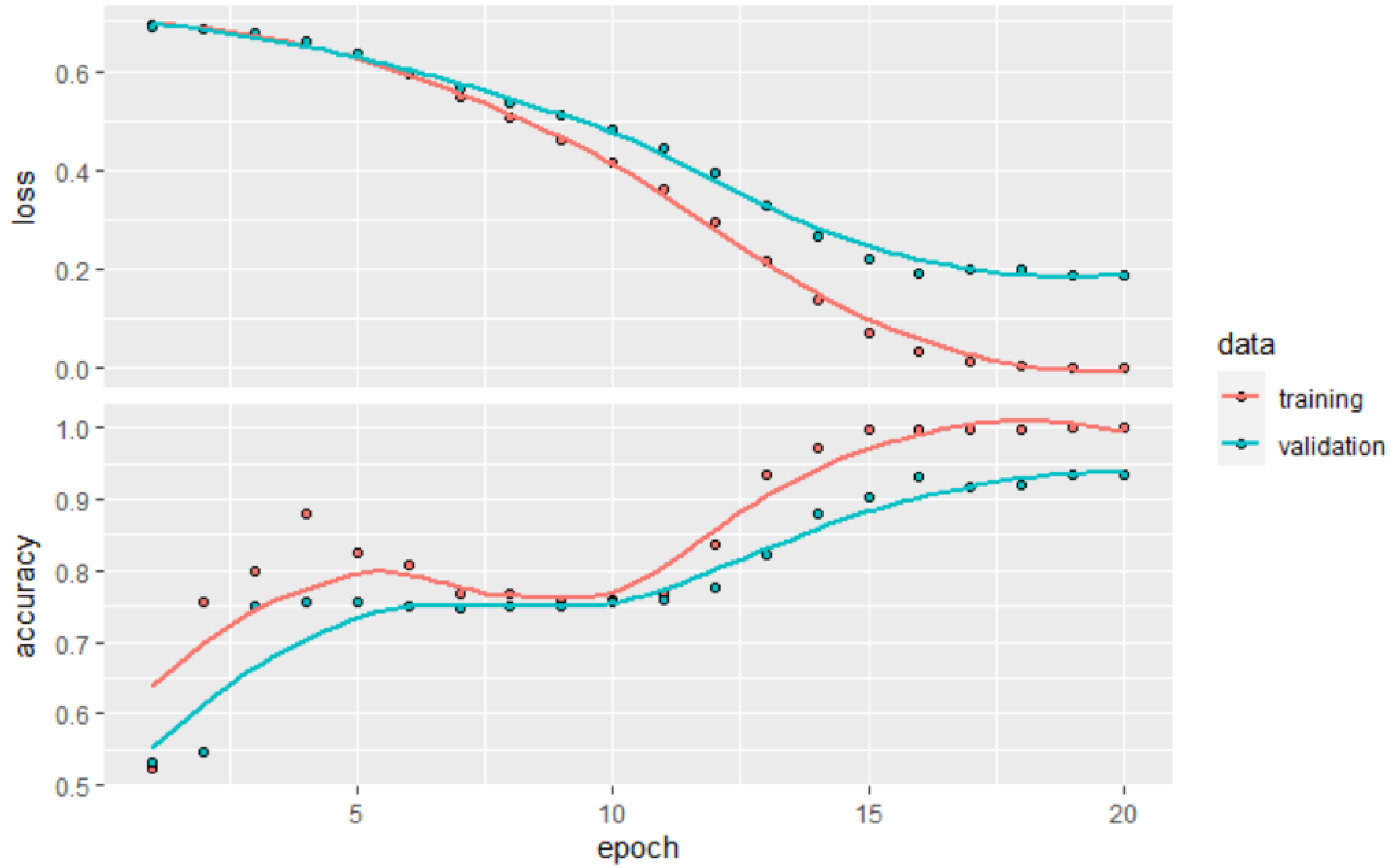
Combines text using a slider of fixed length (kernel), which results in the shortening of the sequence



Source: Supervised Machine Learning for Text Analysis in R

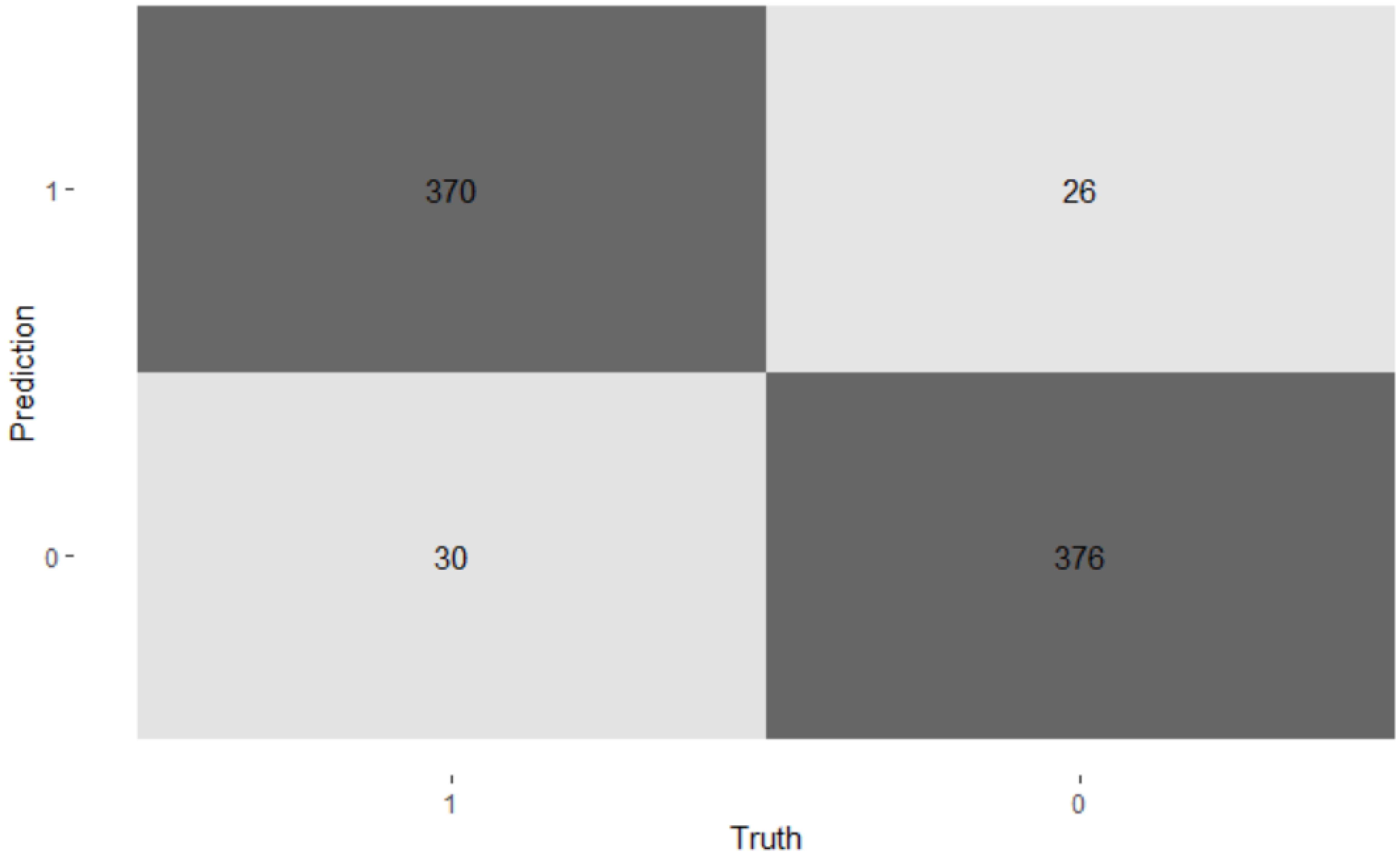
ACCURACY AND LOSS METRICS OVER 20 EPOCHS

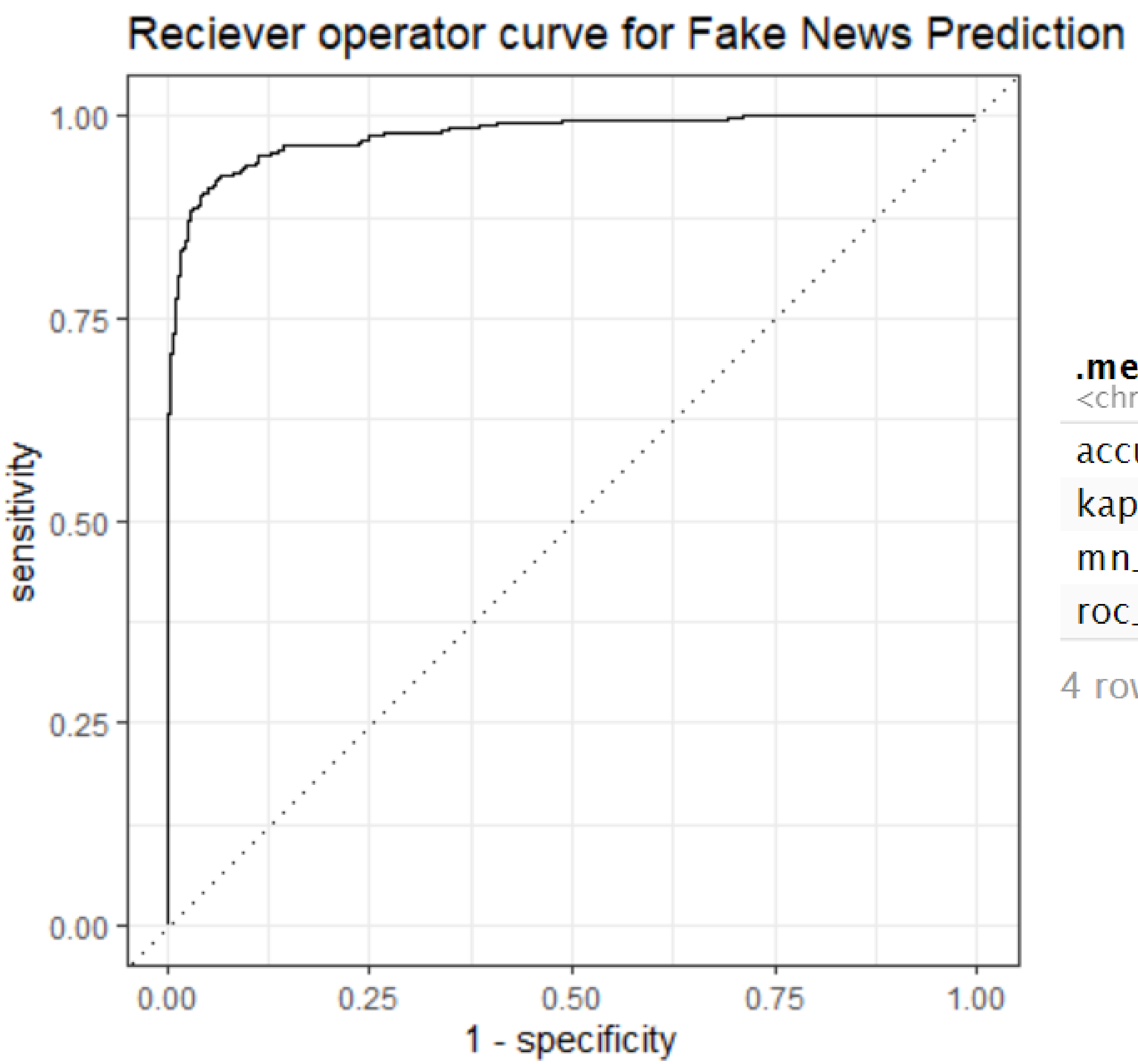
CNN



CONFUSION MATRIX ON OUR INITIAL VALILDATION SET

LSTM





ROC CURVE

.metric <chr>	.estimator <chr>	.estimate <dbl>
accuracy	binary	0.9301746
kap	binary	0.8603448
mn_log_loss	binary	0.2026349
roc_auc	binary	0.9770771

4 rows

COMPARISON OF THE MODELS

	Test accuracy	Kappa value	Loss value	Training time
DNN	90.02494%	0.8002640	0.2376600	2 seconds
LSTM	91.27182%	0.8254310	0.2788496	180 seconds
CNN	93.01746%	0.8603448	0.2026349	20 seconds

Figure 16. Accuracy metrics and Training time for all three neural network models.

CNN - Best, Most Accurate Model

CONCLUSIONS

- Trained multiple types of Neural Networks
- CNN with best balance of time and accuracy
- Accuracy improved with Byte Pair Encoding
- Overfitting reduced through dropout
- Improved accuracy on validation set over more epochs

COMPARED TO SOURCE PAPER

Model	Val. Accuracy	Loss	Val. Loss	Pretraining Time	Finetuning Time
Siamese Networks	77.42%	0.5601	0.5329	N/A	4m per epoch
BERT	87.47%	0.4655	0.4419	66 hours	2m per epoch
GPT-2	90.99%	0.2172	0.1826	78 hours	4m per epoch
ULMFiT	91.59%	0.3750	0.1972	11 hours	2m per epoch
ULMFiT (no LM Finetuning)	78.11%	0.5512	0.5409	11 hours	2m per epoch
BERT + Multitasking	91.20%	0.3155	0.3023	66 hours	4m per epoch
GPT-2 + Multitasking	96.28%	0.2609	0.2197	78 hours	5m per epoch

Comparable performance with much less training time to Cruz et al. (2020)

RECOMMENDATIONS

- Consider improving our model (stopwords and tuning hyperparameters like kernel size)
- Create our own Fake News dataset to test our model
- Examine other categories of news articles (Opinion, satire, sports, showbiz)
- Consider news articles written in other Filipino languages

REFERENCES

Cruz, J. C. B., Tan, J. A., and Cheng, C. (2020, July 1). Localization of Fake News Detection via Multitask Transfer Learning. Proceeding of the 12th Conference on Language Resources and Evaluation (LREC 2020), 2596-2604. https://aclanthology.org/2020.lrec-1.316.pdf

Dobilas, S. (2022, March 5). LSTM Recurrent Neural Networks — How to Teach a Network to Remember the Past. Medium. https://towardsdatascience.com/lstm-recurrent-neural-networks-how-to-teach-a-network-to-remember-the-past-55e54c2ff22e

Ghelani, S. (2019, June 2). *Text Classification — RNN's or CNN's?* Medium; Towards Data Science. https://towardsdatascience.com/text-classification-rnns-or-cnn-s-98c86a0dd361

Kumar, A. (2020, October 28). Keras - Categorical Cross Entropy Loss Function. Data Analytics. https://vitalflux.com/keras-categorical-cross-entropy-loss-function/

Lalu, G. P. (2022, October 11). 'fake news' a problem in ph? 9 in 10 Filipinos agree, says pulse Asia. INQUIRER.net. Retrieved November 27, 2022, from https://newsinfo.inquirer.net/1678248/fake-news-a-problem-in-ph-9-in-10-filipinos-agree-says-pulse-asia?fbclid=IwAR0eh-qMvTIyjzQ_9o-FUaKqWF531aTKfhgynfqiat6a0hh7_qpvfMm6ADs

REFERENCES

Siar, S. (2021, August). Fake news, its dangers, and how we can fight it. Retrieved November 27, 2022, from https://pidswebs.pids.gov.ph/CDN/PUBLICATIONS/pidspn2106.pdf

Silge, E. H. and J. (n.d.). Chapter 10 Convolutional neural networks | Supervised Machine Learning for Text Analysis in R. In smltar.com. Retrieved November 27, 2022, from https://smltar.com/dlcnn.html

De Paor, S., & Heravi, B. (2020). Information literacy and fake news: How the field of librarianship can help combat the epidemic of fake news. The Journal of Academic Librarianship, 46(5), 102218. https://doi.org/10.1016/j.acalib.2020.102218

Hvitfeldt, E., & Silge, J. (2022). Supervised machine learning for text analysis in R. CRC Press.