# ULMFIT FOR POLISH HATE SPEECH DETECTION

*Piotr Czapla, Marcin Kardas // n-waves*
*Jeremy Howard, Sylvain Gugger // fast.ai*

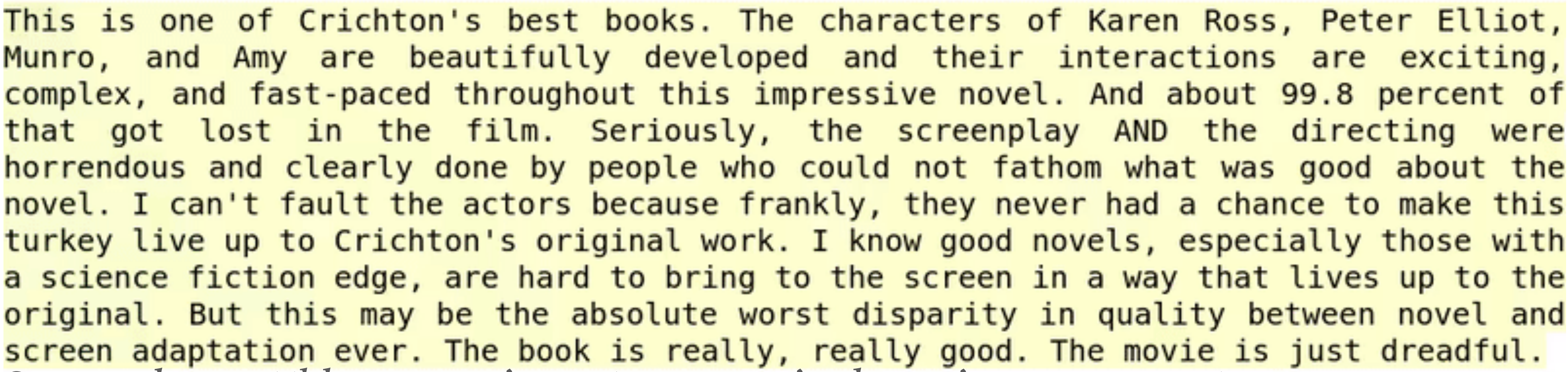
*AI & NLP workshop day*
*May 31, 2019*

# TRANSFER LEARNING

n-waves

# TRANSFER LEARNING

- Char level lang model learns sentiment



This is one of Crichton's best books. The characters of Karen Ross, Peter Elliot, Munro, and Amy are beautifully developed and their interactions are exciting, complex, and fast-paced throughout this impressive novel. And about 99.8 percent of that got lost in the film. Seriously, the screenplay AND the directing were horrendous and clearly done by people who could not fathom what was good about the novel. I can't fault the actors because frankly, they never had a chance to make this turkey live up to Crichton's original work. I know good novels, especially those with a science fiction edge, are hard to bring to the screen in a way that lives up to the original. But this may be the absolute worst disparity in quality between novel and screen adaptation ever. The book is really, really good. The movie is just dreadful.

*Source: https://blog.openai.com/unsupervised-sentiment-neuron/*

- GPT-2 unsupervised
    - Question answering
    - Translation
    - Summarisation

# TRANSFER LEARNING & ULMFIT

- Classification: ~20% error reduced
- Quick pre-training = fast experiments
  - (4h Reddit lub 18h Wikipedia)
  - 2 min train on PolEval 2019 hate speech dataset
- Better than LASER and BERT on MLDoc & Cross-Lingual Sentiment dataset (in review)

# WHAT'S ULMFIT

*+ our adaptation for Polish*

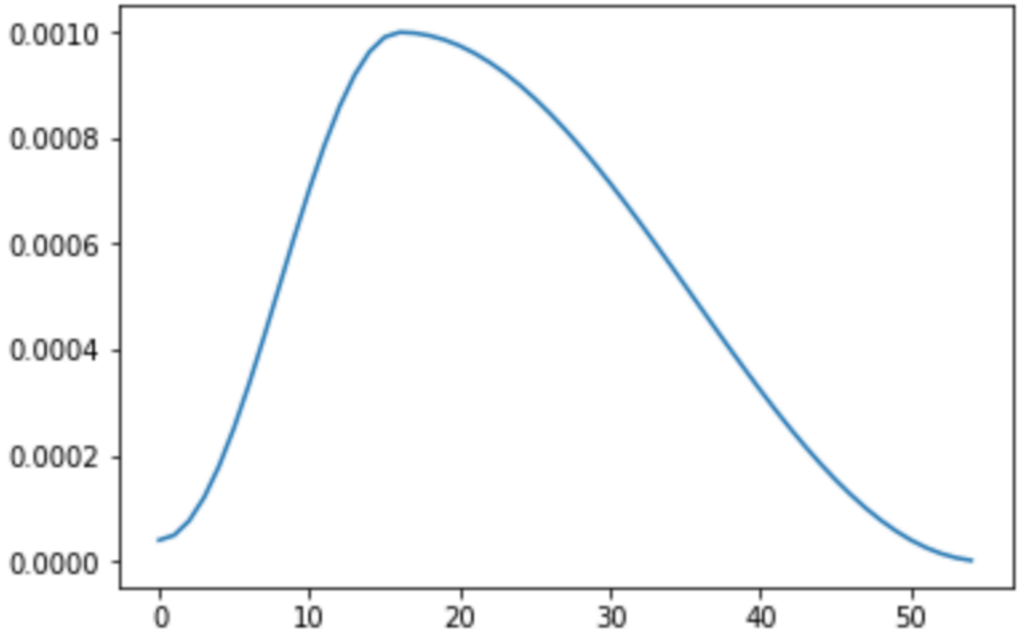
n-waves

# GOOD LM

- Good Language Model based on 3 layer LSTM with tweaks by Stephen Merity
  - Weight decay, Temporal Activation Regularization, Activation Regularization
  - 5 different dropout layers
  - Gradient clipping

# ULMFIT - TRANSFER LEARNING

- Fast.ai
  - 1 cycle training policy
  - discriminative learning rate
  - fast.ai preprocessing
  - **Well tuned classification heac**



n-waves

# ULMFIT - HATE SPEECH

- Sentence Piece - to handle polish
- Small vocabulary size 25k
- 4 layers instead of 3
- Well tuned learning rate
- Small framework to iterate fast
- **Pre-training on Reddit (+5 pp.)**
- **Weighted cross-entropy (+10 pp.)**

# HOW TO TRAIN

*To get good results and stay sane*

n-waves

# ABLATION STUDIES - PRE-TRAINING MATTERS

- Use close enough corpus for pre-training

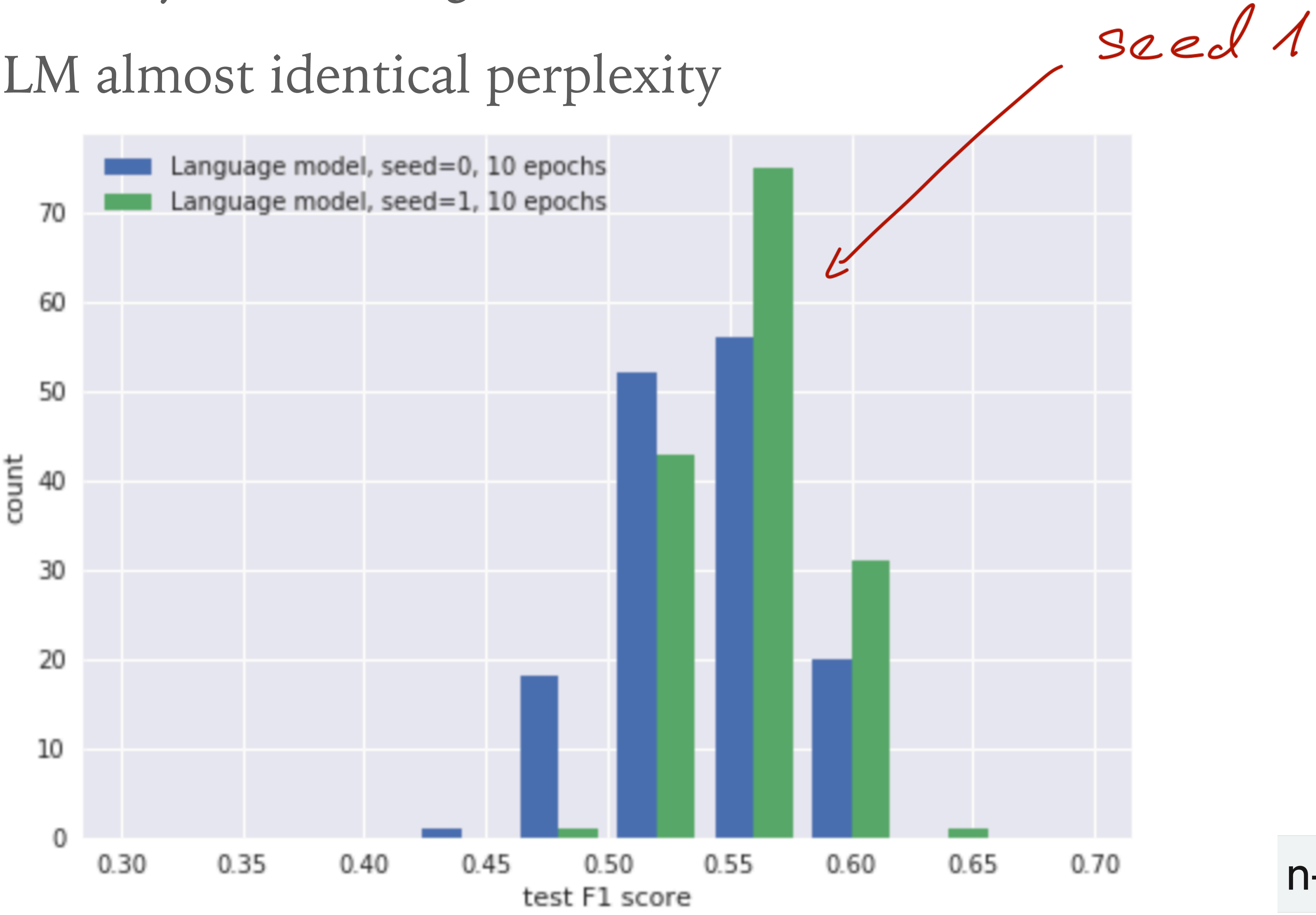
| | 50% | 75% |
|---|---|---|
| Wiki | 52,02 | 54,19 |
| Wiki fine tune | 53,77 | 55,88 |
| Reddit | 57,49 | 58,14 |

↑ no fine tuning

| | |
|---|---|
| *Twitter* | *To żaden problem. Amerykanie to …* |
| *Reddit* | *Byłem kiedyś u fryzjera i obciął mi włosy.* |
| *Wikipedia* | *Borland ( w latach 1998-2001 pod nazwą Inprise ) – amerykańskie* |

# ABLATION STUDIES - INITIALISATION MATTERS

- Initially drawn weights matter
- LM almost identical perplexity

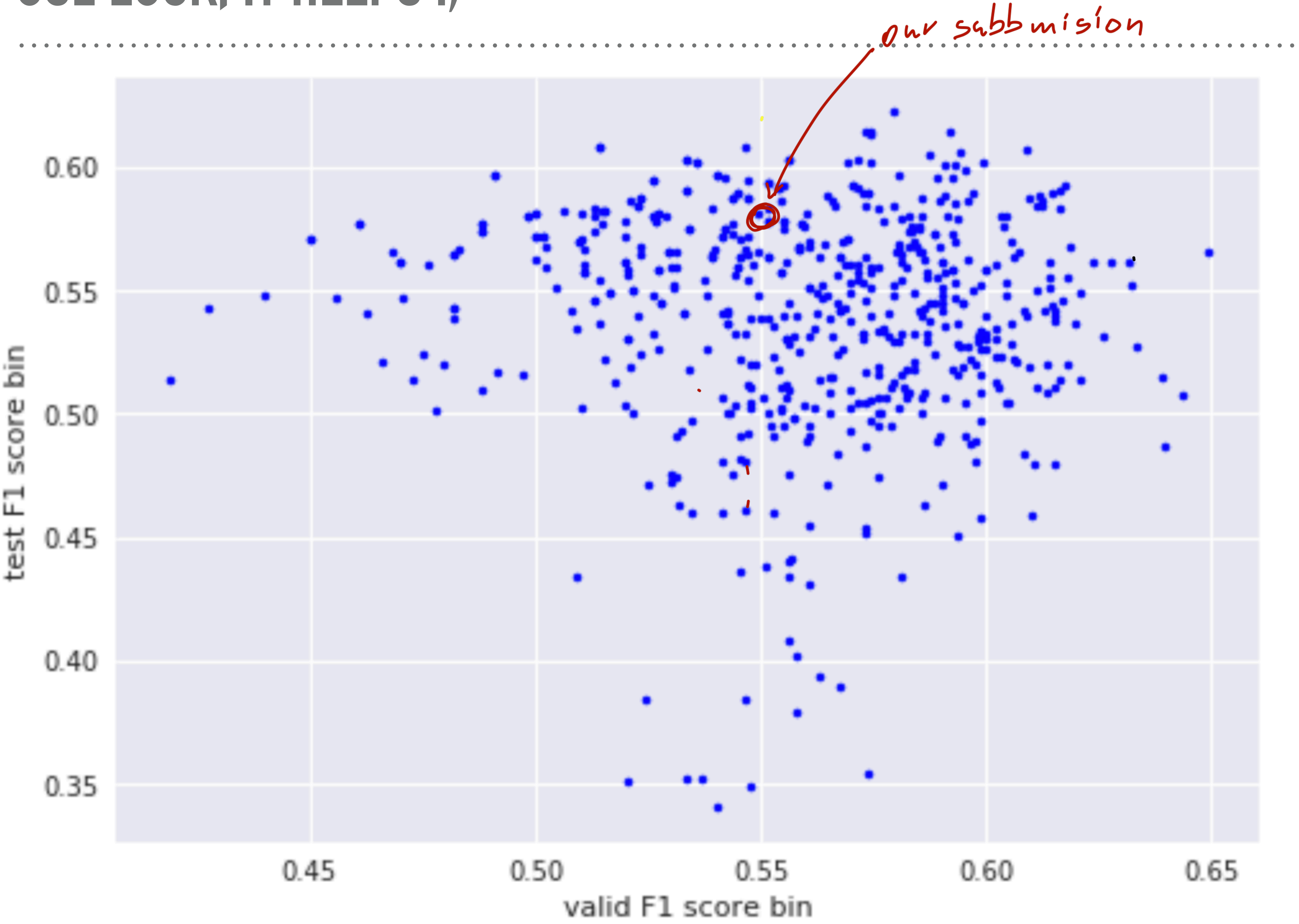


n-waves

# ABLATION STUDIES - FINE TUNE CLASSIFIERS

- Weighted cross entropy increases F1
    - 54% vs 44%, mean of 38 models
- Early stopping and additional preprocessing doesn't matter
- Keeping random seeds fixed helps you stay sane
- Train a lot of models before you draw conclusions (~1000)

# USE LUCK, IT HELPS :)



# CONTRIBUTION WANTED

- State of the art techniques applied to: CV, NLP, Tabular, Collab
- Excellent courses and vibrant community
- Focus on applicability and small scale learning

fast.ai

- Brings NLP breakthroughs to local languages
- Commercial Transfer learning
- Open source Research
- Plenty of room for contribution

n-waves

# THANK YOU