

# Named Entity and Relation Extraction From Wet Lab Protocols



**Jeniya Tabassum**



Sydney Lee



Alan Ritter



Wei Xu

# ProtocolsIO



**protocols.io**

 nCoV-2019 sequencing protocol  
Josh Quick<sup>1</sup>  
<sup>1</sup>University of Birmingham  
11 Works for me dx.doi.org/10.17504/protocols.io.bdp7i5rn  
ARTIC Coronavirus Method Development Community

Version 2 ▾ Apr 09, 2020

1 Mix the following components in an 0.2mL 8-strip tube;

Component	Volume
50µM random hexamers	1 µl
10mM dNTPs mix (10mM each)	1 µl

 Viral RNA input from a clinical sample should be between Ct 18-35. If Ct is between 12-15, then dilute the sample 100-fold in water, if between 15-18 then dilute 10-fold in water. This will reduce the likelihood of PCR-inhibition.

2 Gently mix by pipetting and pulse spin the tube to collect liquid at the bottom of the tube.

3 Incubate the reaction as follows:

65 °C for 00:05:00  
Place on ice for 00:01:00

4 Add the following to the annealed template RNA:

Component	Volume
SSIV Buffer	4 µl
100mM DTT	1 µl
RNaseOUT RNase Inhibitor	1 µl
SSIV Reverse Transcriptase	1 µl
Total	20 µl

# ProtocolsIO



**protocols.io**

**Reagent**      **Method**

**nCoV-2019 sequencing protocol**

<sup>1</sup>University of Birmingham

11 Works for me    dx.doi.org/10.17504/protocols.io.bdp7i5rn

ARTIC    Coronavirus Method Development Community

Mix the following components in an 0.2mL 8-strip tube;

Viral RNA input from a clinical sample should be between Ct 18-35.

If Ct is between 12-15, then dilute the sample 100-fold in water, if between 15-18 then dilute 10-fold in water.

This will reduce the likelihood of PCR-inhibition.

A mastermix should be made up in the mastermix cabinet and aliquoted into PCR strip tubes.

Tubes should be wiped down when entering and leaving the mastermix cabinet.

Gently mix by pipetting and pulse spin the tube to collect liquid at the bottom of the tube.

Incubate the reaction as follows:

65 °C for 00:05:00

Place on ice for 00:01:00

Add the following to the annealed template RNA:

1

# ProtocolsIO

Machine Readable Text



Reagent Method  
nCoV-2019 sequencing protocol

1 University of Birmingham

11 Works for me dx.doi.org/10.17504/protocols.io.bdp7i5rn

ARTIC Coronavirus Method Development Community

Version 2 ▾ Apr 09, 2020

Mix the following components in an 0.2mL 8-strip tube;  
Viral RNA input from a clinical sample should be between Ct 18-35.

If Ct is between 12-15, then dilute the sample 100-fold in water, if between 15-18 then dilute 10-fold in water.

This will reduce the likelihood of PCR-inhibition.

A mastermix should be made up in the mastermix cabinet and aliquoted into PCR strip tubes.

Tubes should be wiped down when entering and leaving the mastermix cabinet.

Gently mix by pipetting and pulse spin the tube to collect liquid at the bottom of the tube.

Incubate the reaction as follows:

65 °C for 00:05:00

Place on ice for 00:01:00

Add the following to the annealed template RNA:

# ProtocolsIO

Entity & Relation Extraction

The screenshot shows a protocol titled "nCoV-2019 sequencing protocol" by University of Birmingham. It's version 2, published on Apr 09, 2020, and has 11 works for it. The DOI is dx.doi.org/10.17504/protocols.io.bdp7i5rn. The protocol text is annotated with entities like Reagent, Method, Site, Amount, Measure, Location, Measure Type, GM, Ct, Rgt, num, and Action, connected by arrows indicating relationships such as "Acts on", "Meronym", "Of Type", "Setting", "Site", "Action", "Measure Type-Link", and "Mod".

Reagent Method  
nCoV-2019 sequencing protocol  
1University of Birmingham  
11 Works for me dx.doi.org/10.17504/protocols.io.bdp7i5rn  
ARTIC Coronavirus Method Development Community

Mix the following components in an 0.2mL 8-strip tube;  
Viral RNA input from a clinical sample should be between Ct 18-35.  
is between 12-15, then dilute the sample 100-fold in water, if between 15-18 then dilute 10-fold in water.  
This will reduce the likelihood of PCR-inhibition.  
A mastermix should be made up in the mastermix cabinet and aliquoted into PCR strip tubes.  
Tubes should be wiped down when entering and leaving the mastermix cabinet.  
Gently mix by pipetting and pulse spin the tube to collect liquid at the bottom of the tube.  
Incubate the reaction as follows:  
65 °C for 00:05:00  
Place on ice for 00:01:00  
Add the following to the annealed template RNA:

Machine Readable Text



# Annotated ProtocolsIO Corpus

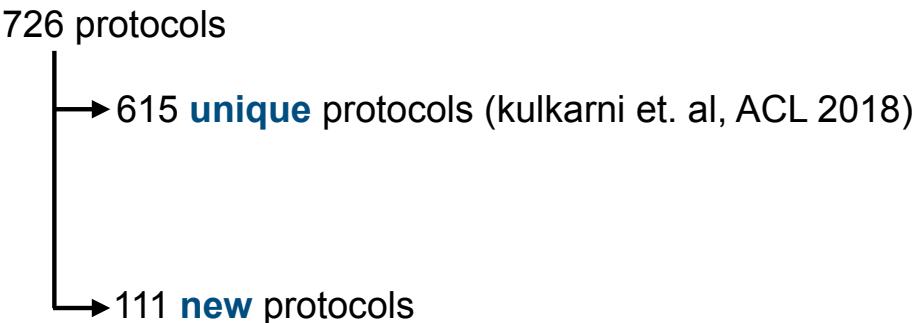
**An Annotated Corpus for Machine Reading of Instructions  
in Wet Lab Protocols**

**Chaitanya Kulkarni, Wei Xu, Alan Ritter, Raghu Machiraju**

Department of Computer Science and Engineering  
Ohio State University



Archive timeline:  
2012 - 2020



# Annotated ProtocolsIO Corpus

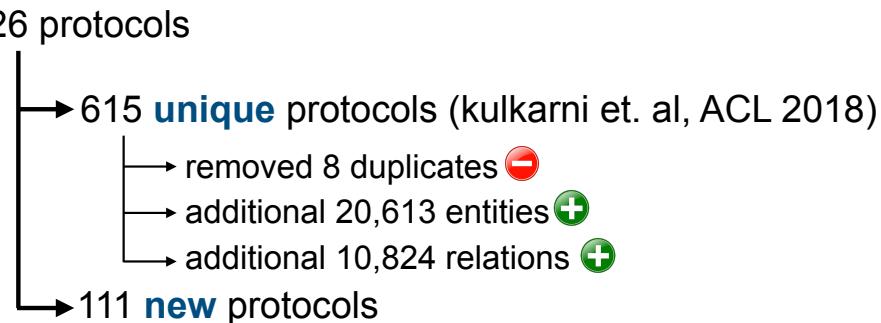
An Annotated Corpus for Machine Reading of Instructions  
in Wet Lab Protocols

Chaitanya Kulkarni, Wei Xu, Alan Ritter, Raghu Machiraju

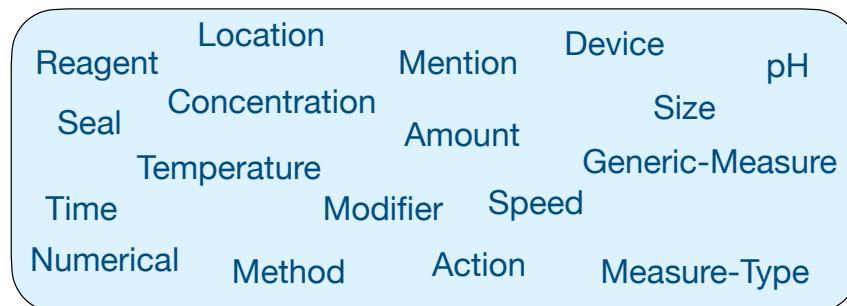
Department of Computer Science and Engineering  
Ohio State University



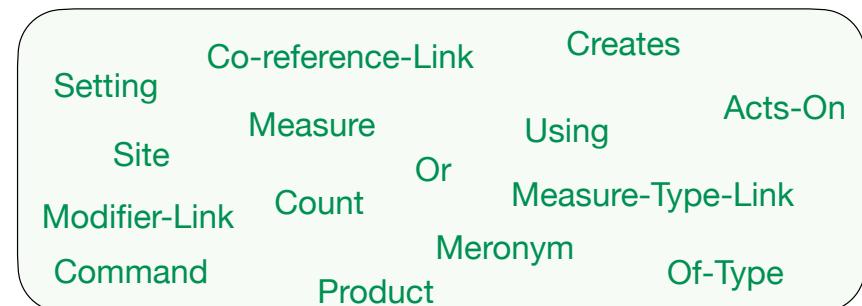
Archive timeline:  
2012 - 2020



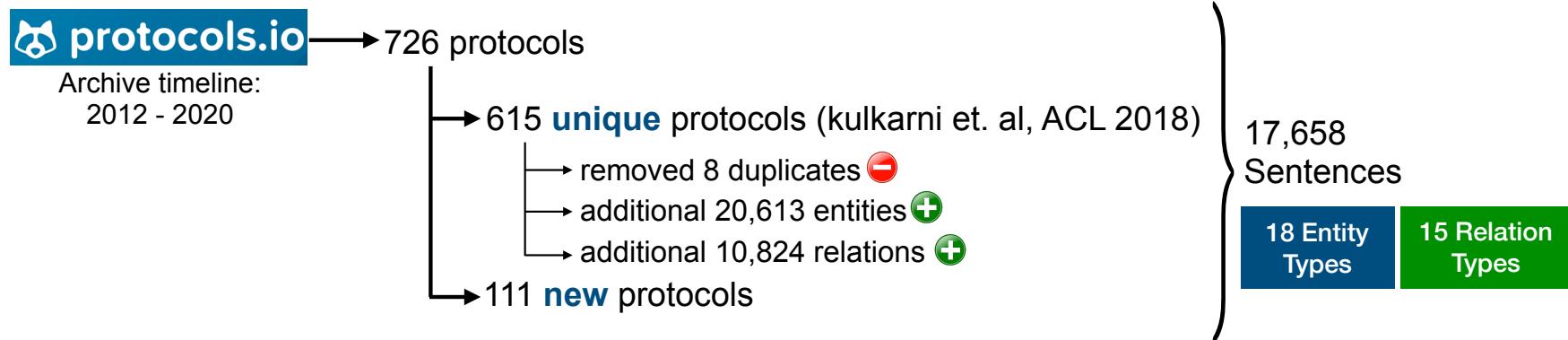
# Annotated ProtocolsIO Corpus



Entity Types



Relation Types



# Shared Task @WNUT-EMNLP-2020



# Shared Task @WNUT-EMNLP-2020

## Sub-Task-1

 Mix the following components in an 0.2mL 8-strip tube;

Reagent  
Amount  
Location

Named Entity Extraction  
(NER)

## Sub-Task-2

 Mix the following components in an 0.2mL 8-strip tube;

Site  
Acts on  
Reagent  
Amount  
Measure  
Location

Relation Extraction  
(RE)

# Shared Task @WNUT-EMNLP-2020

## Sub-Task-1

 Mix the following components in an 0.2mL 8-strip tube;

Reagent      Amount      Location

Named Entity Extraction  
(NER)

## Sub-Task-2

 Mix the following components in an 0.2mL 8-strip tube;

Site      Reagent      Amount      Location

Acts on      Measure

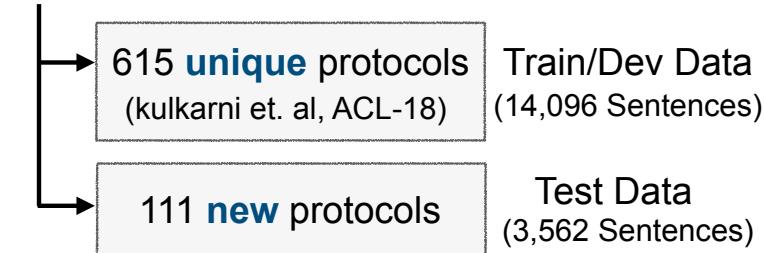
Relation Extraction  
(RE)



protocols.io

Archive timeline:  
2012 - 2020

→ 726 protocols



# Shared Task @WNUT-EMNLP-2020

## Sub-Task-1

 Mix the following components in an 0.2mL 8-strip tube;

Reagent      Amount      Location

Named Entity Extraction  
(NER)

## Sub-Task-2

 Mix the following components in an 0.2mL 8-strip tube;

Site      Reagent      Amount      Location

Acts on      Measure

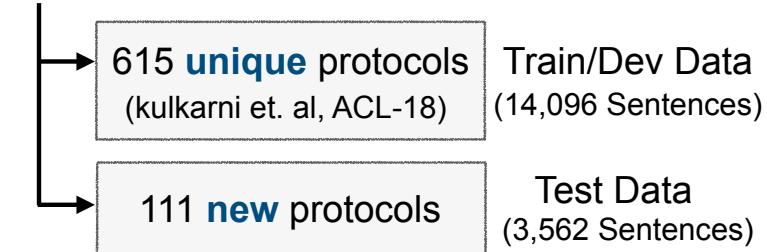
Relation Extraction  
(RE)



**protocols.io**

Archive timeline:  
2012 - 2020

→ 726 protocols



# NER Techniques

Neural Ensemble

BiTeM  
PubCov19

Bert Fine-Tuning

Facny Man  
mahab  
mgsohrab  
SudesnaTCS  
IITKGP

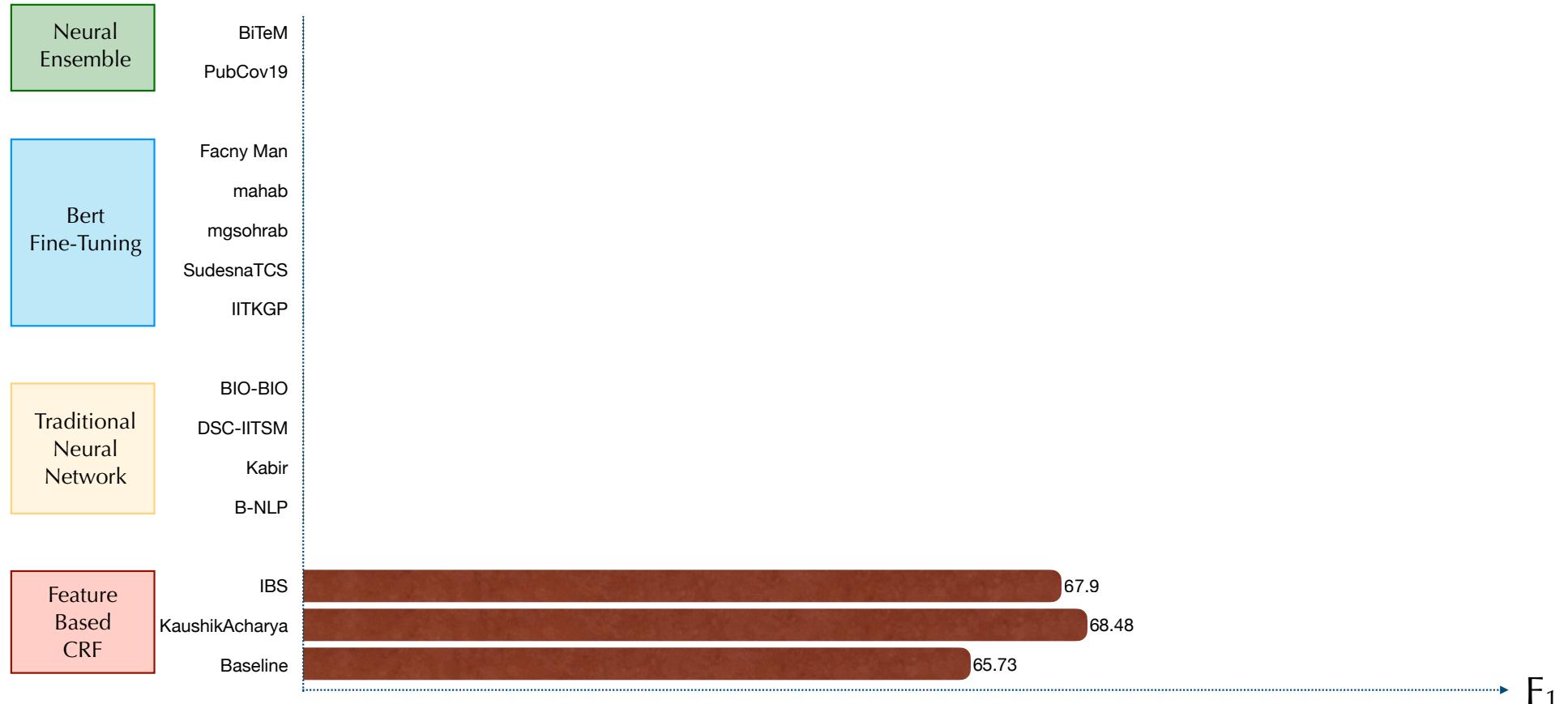
Traditional Neural Network

BIO-BIO  
DSC-IITSM  
Kabir  
B-NLP

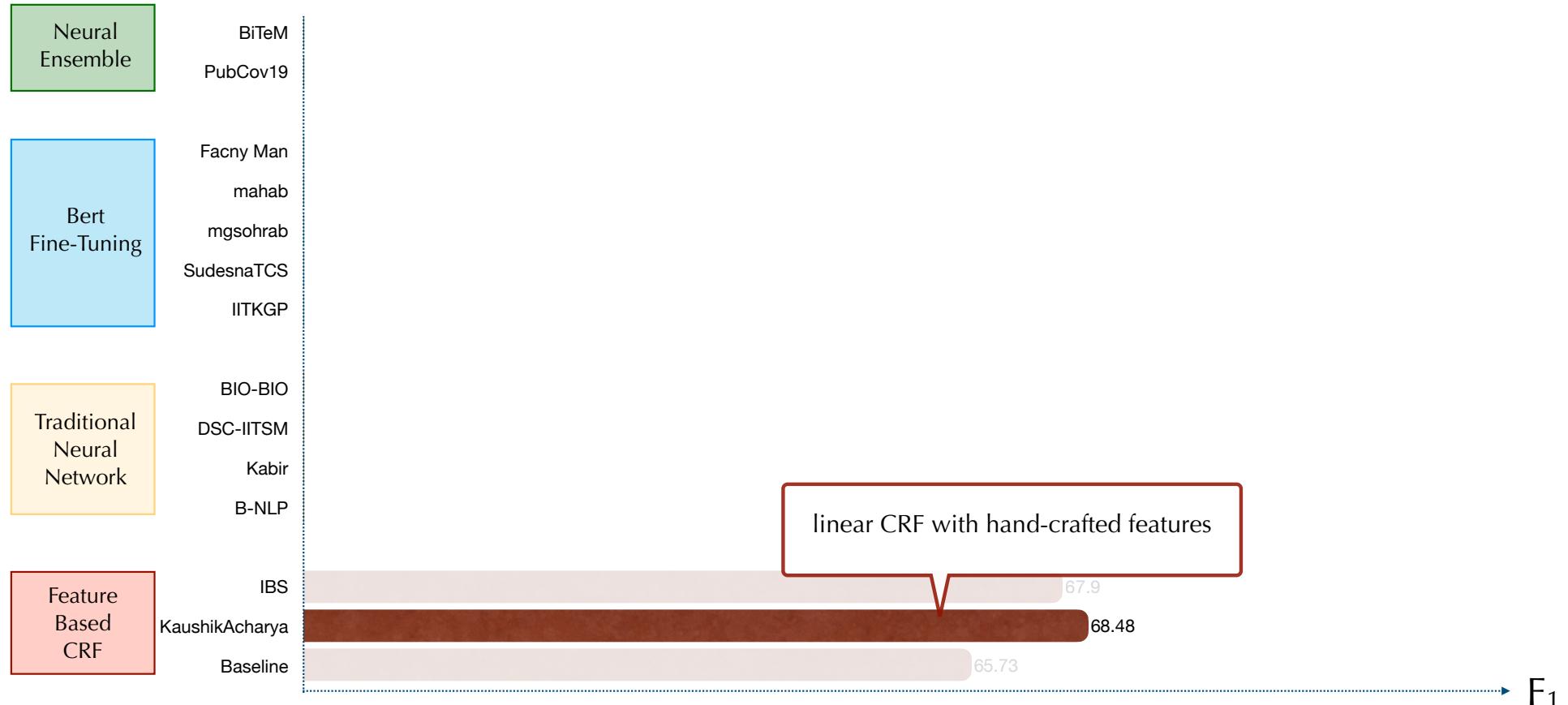
Feature Based CRF

IBS  
KaushikArchaya  
Baseline

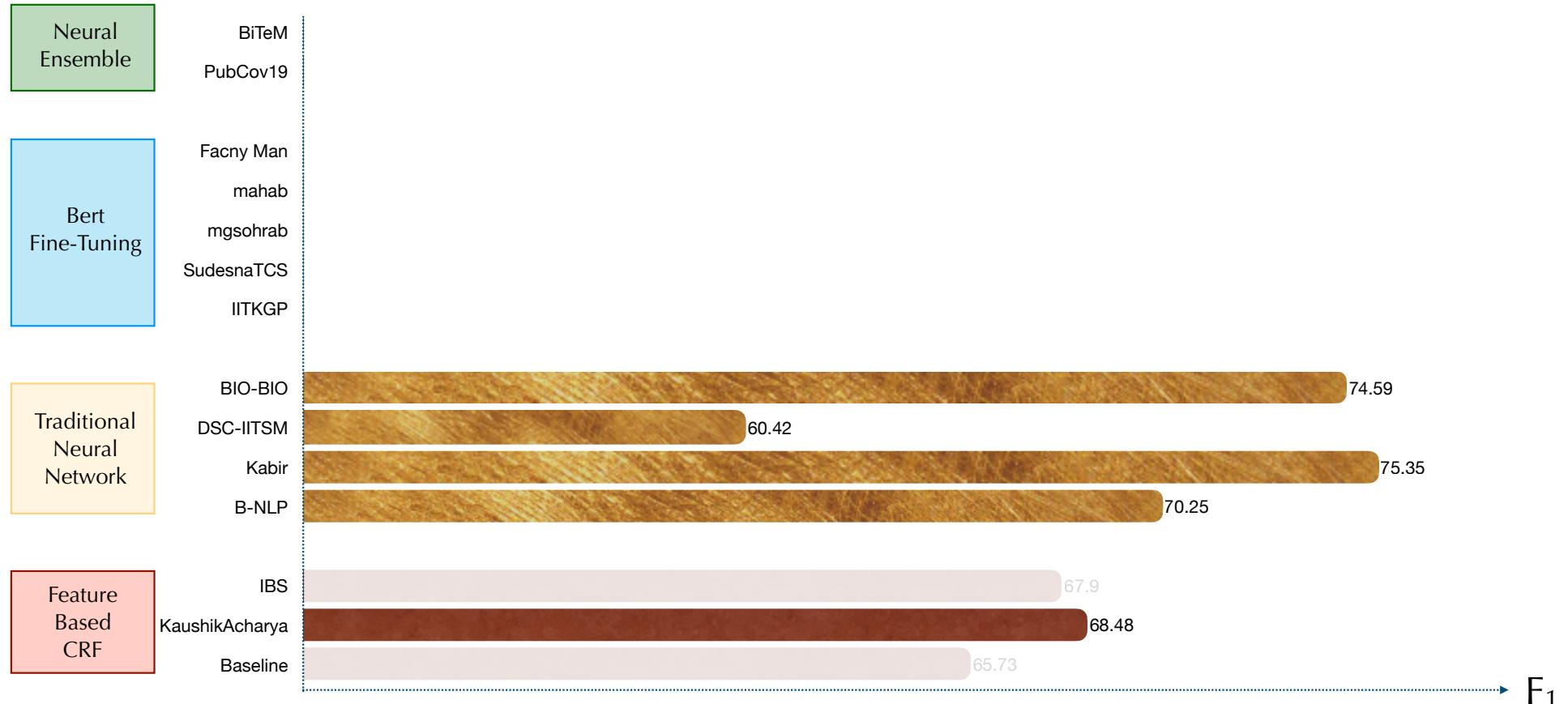
# NER Techniques



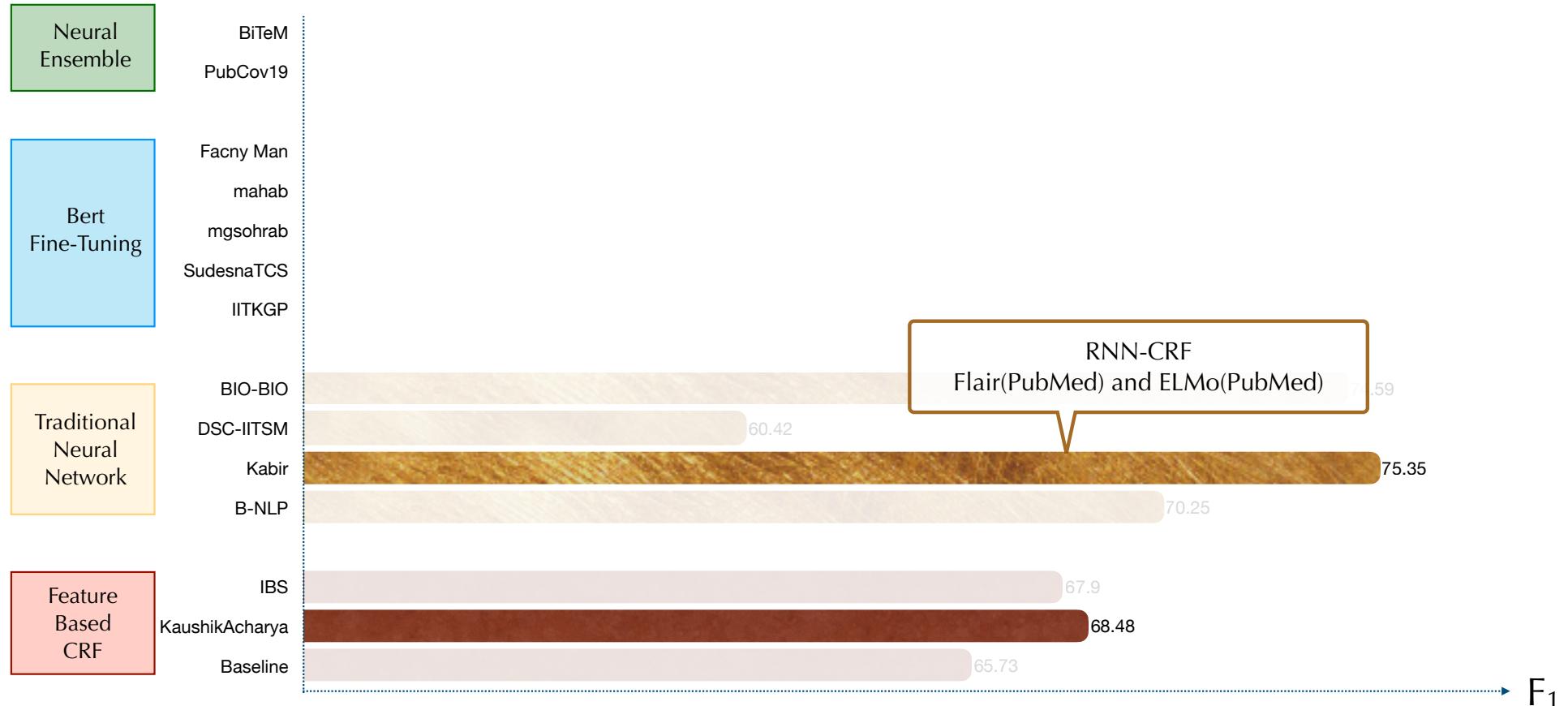
# NER Techniques



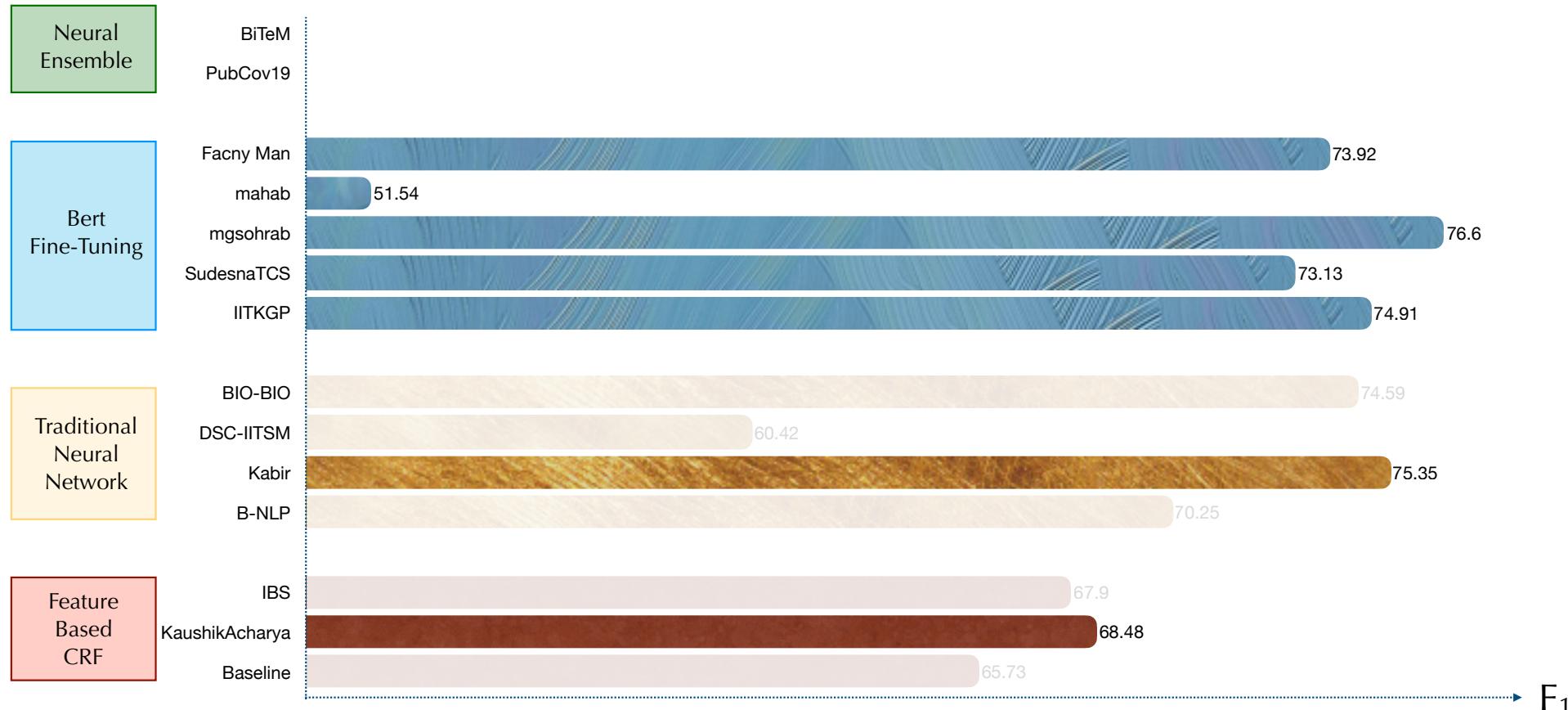
# NER Techniques



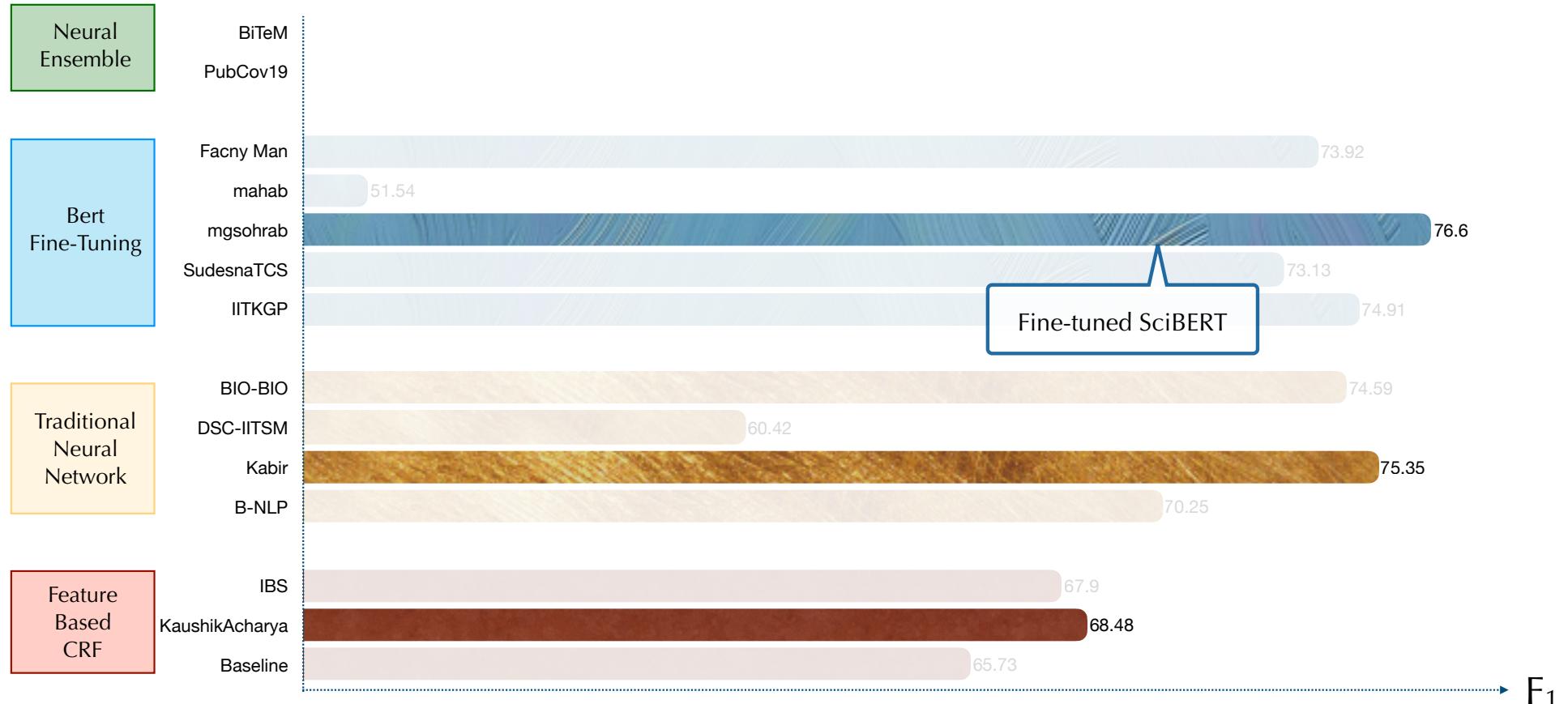
# NER Techniques



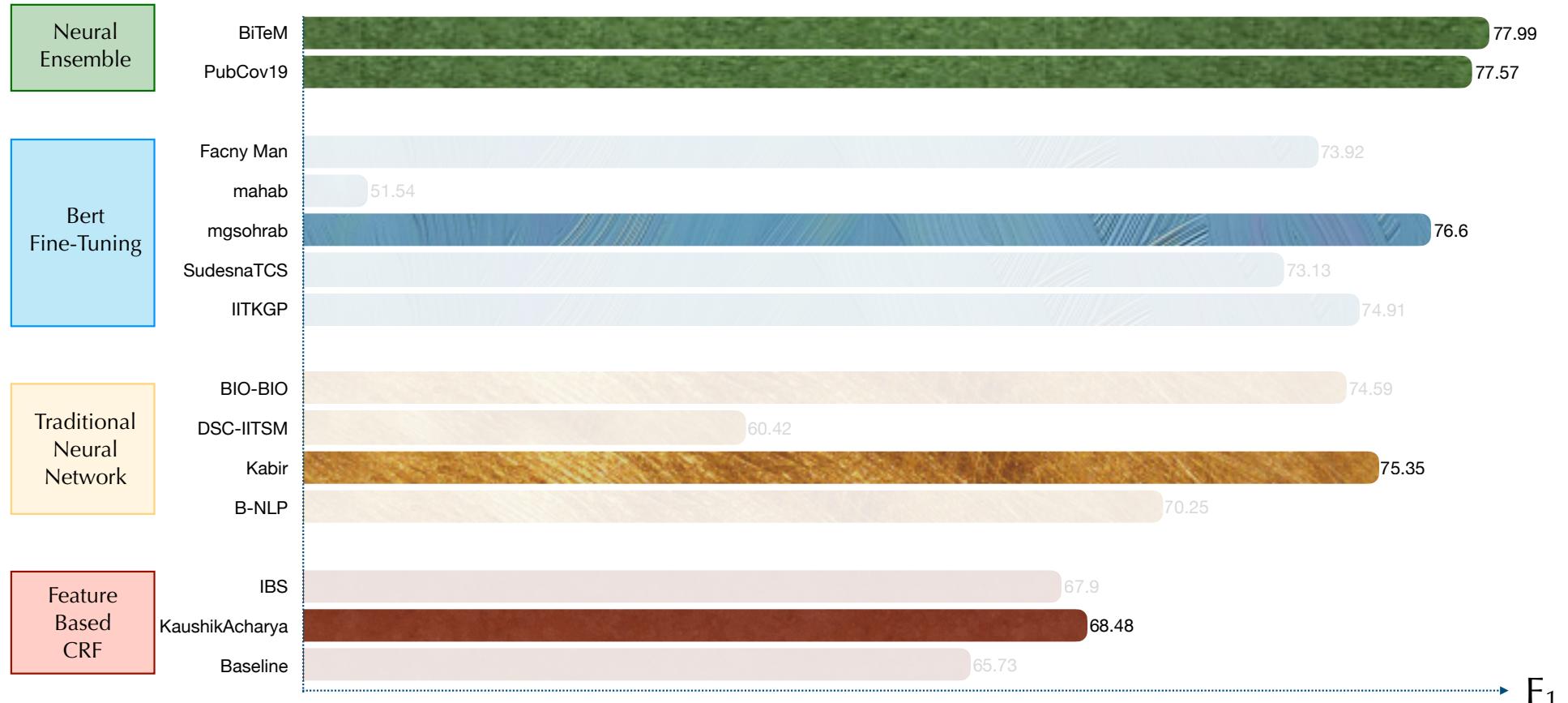
# NER Techniques



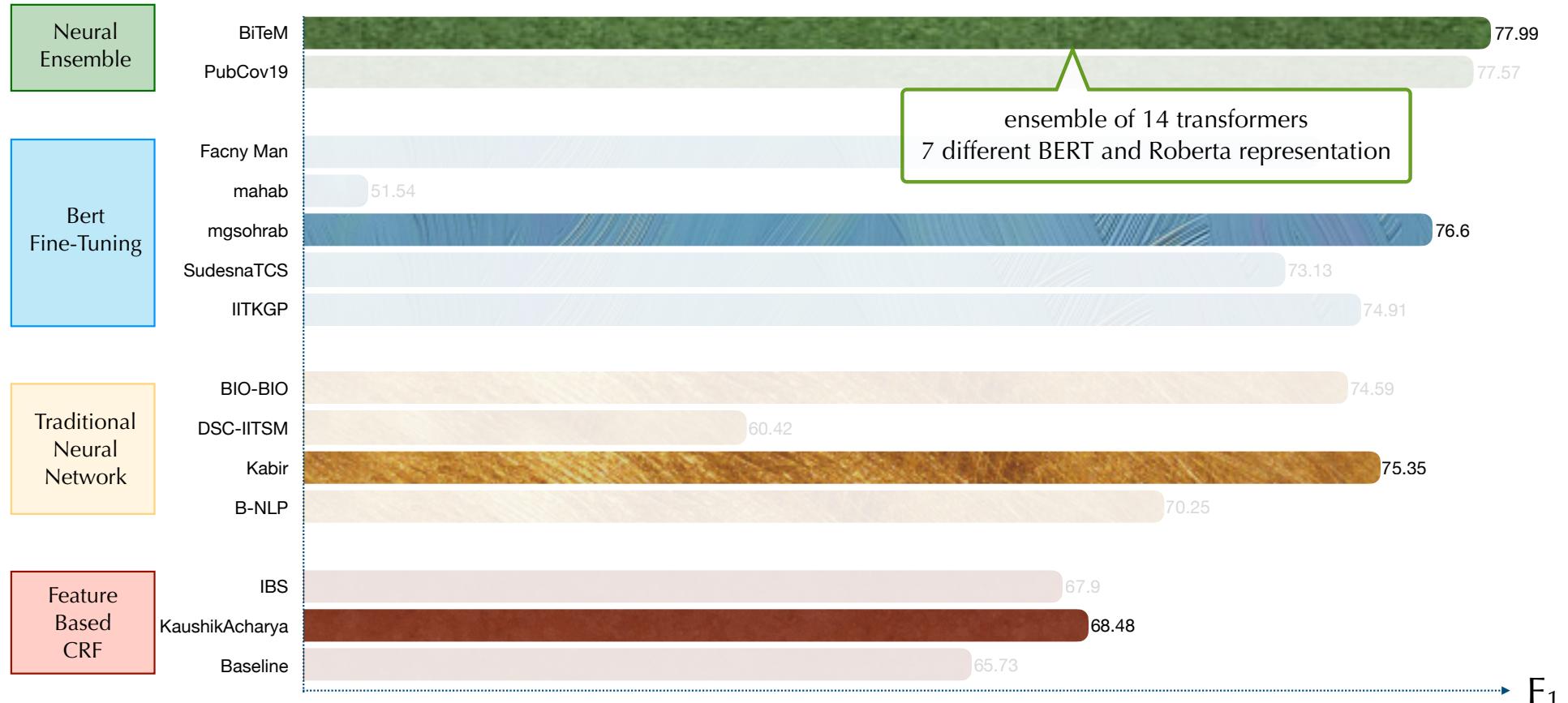
# NER Techniques



# NER Techniques



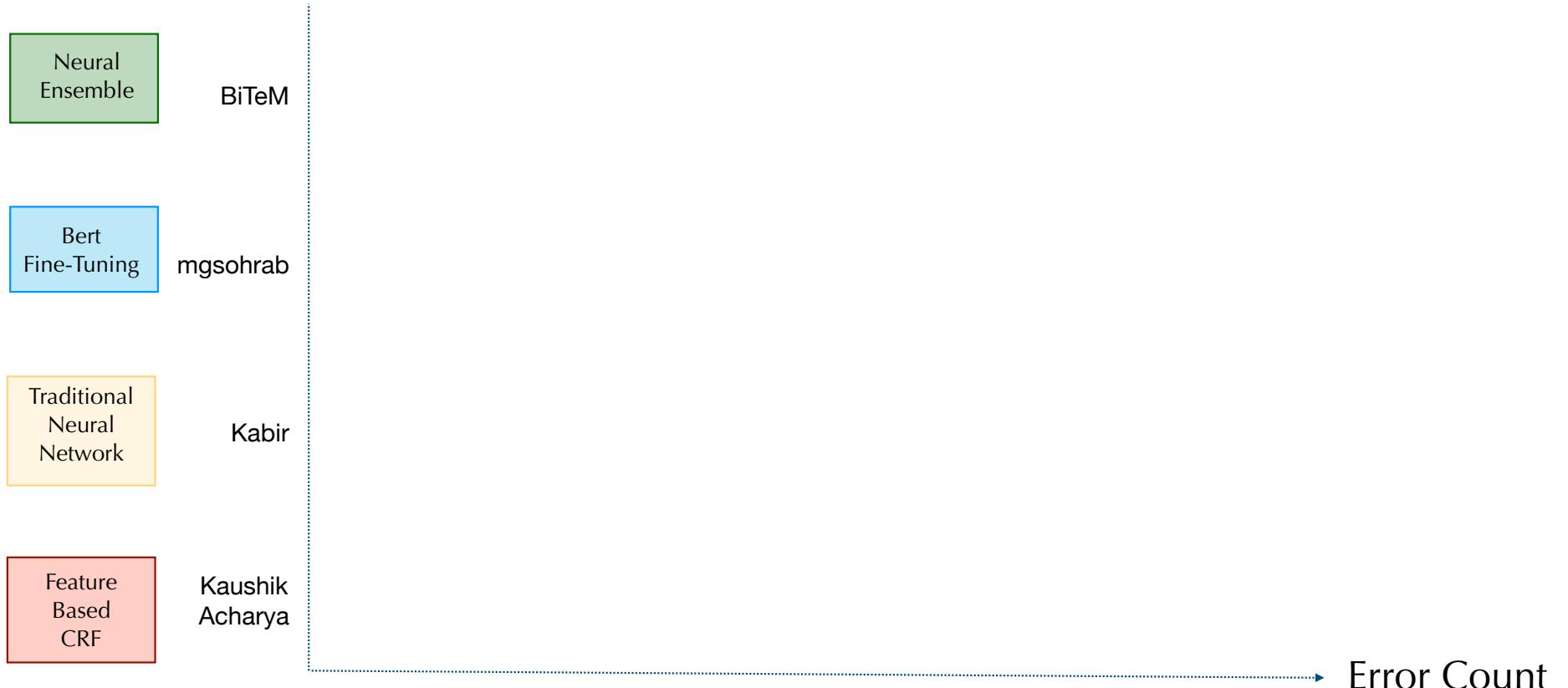
# NER Techniques



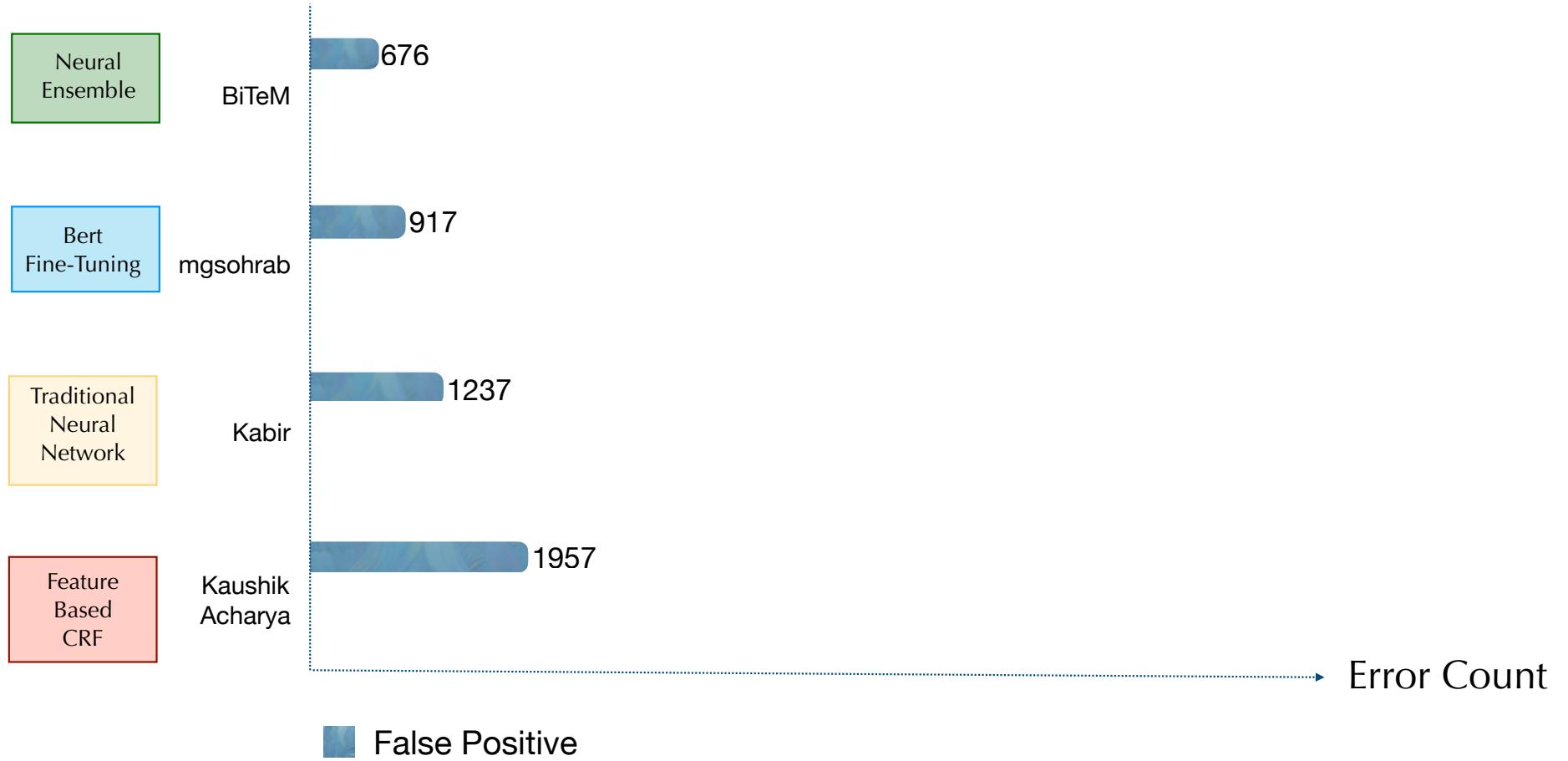
# NER Techniques



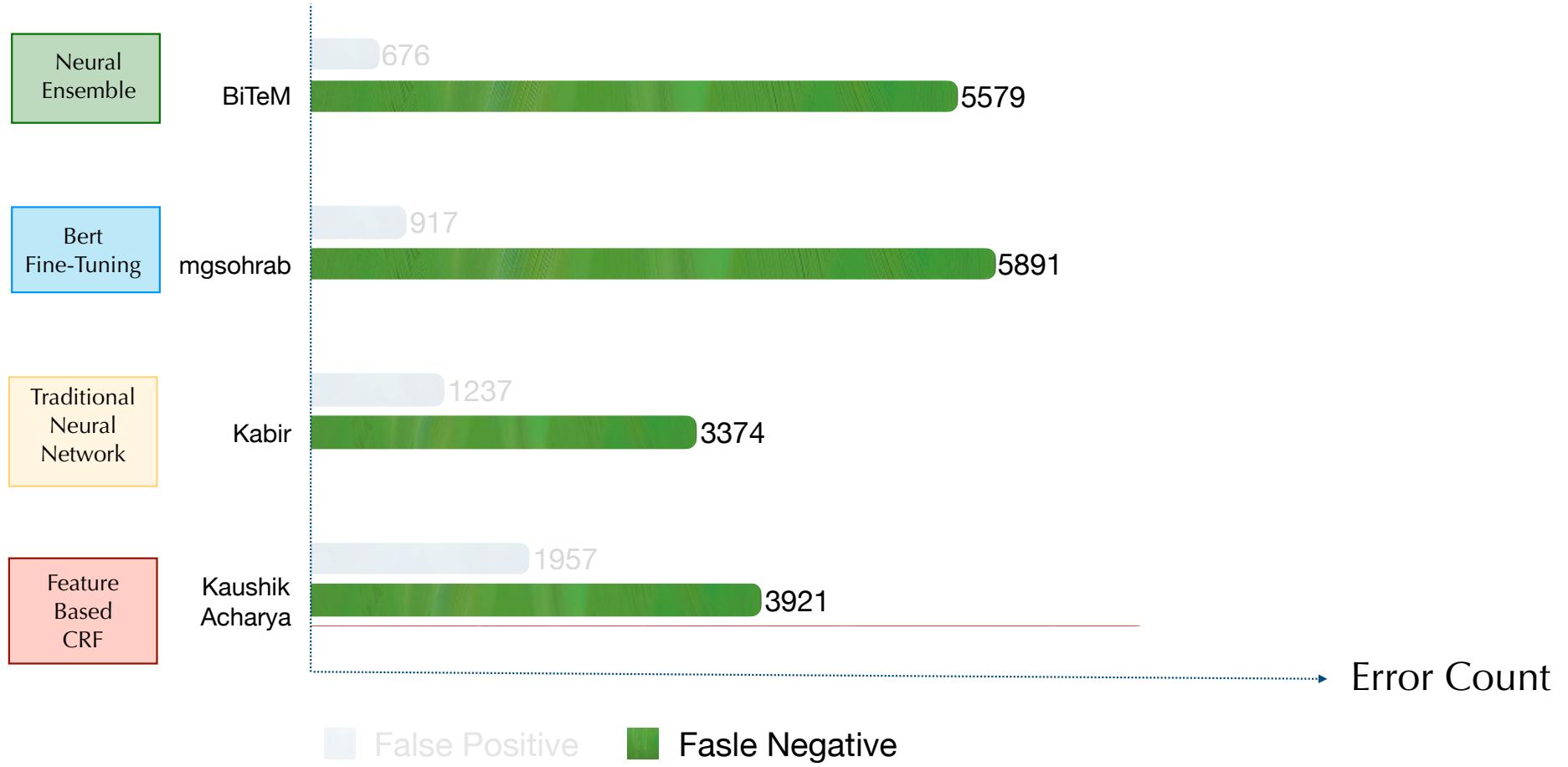
# NER Error Distribution



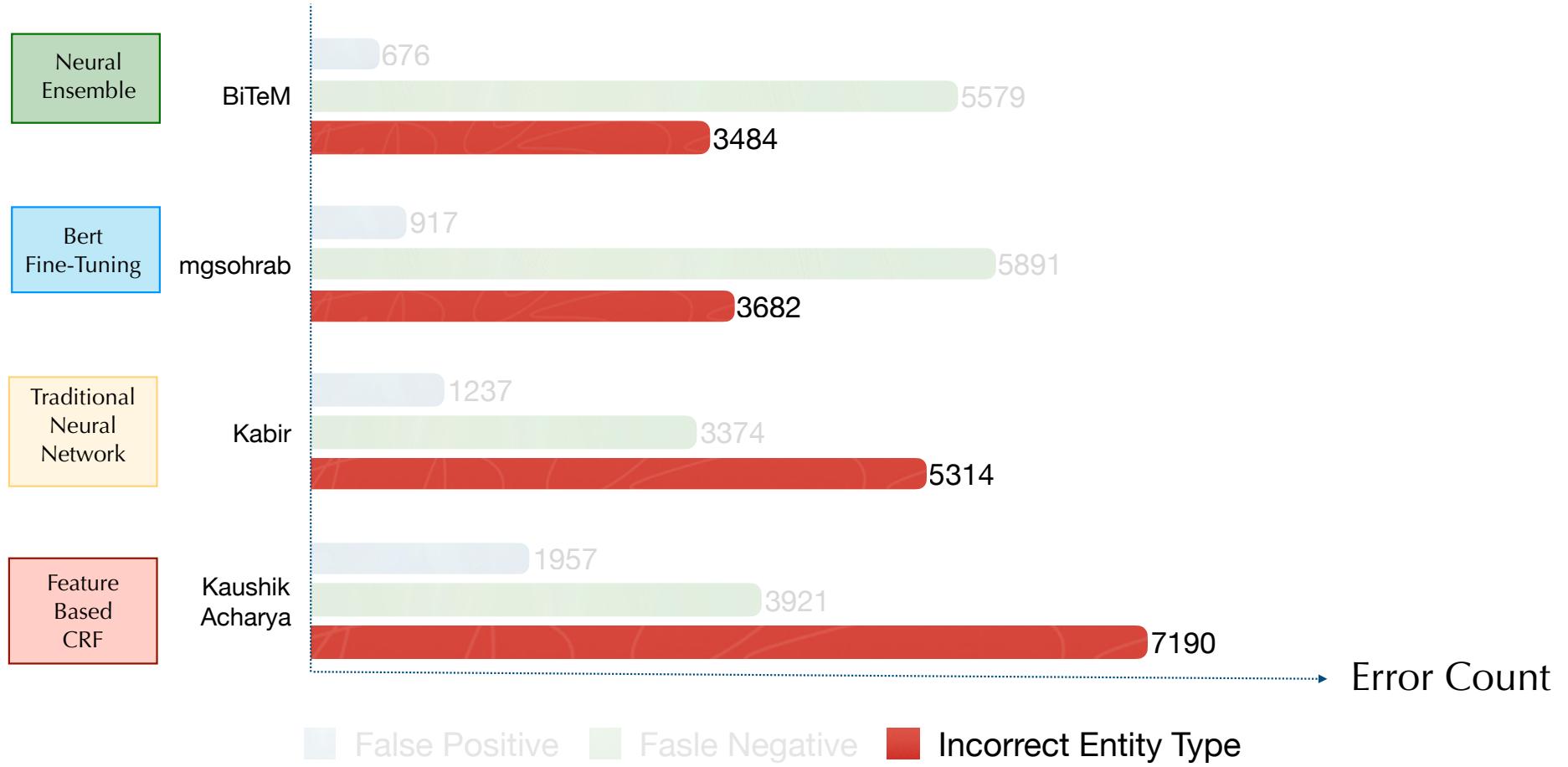
# NER Error Distribution



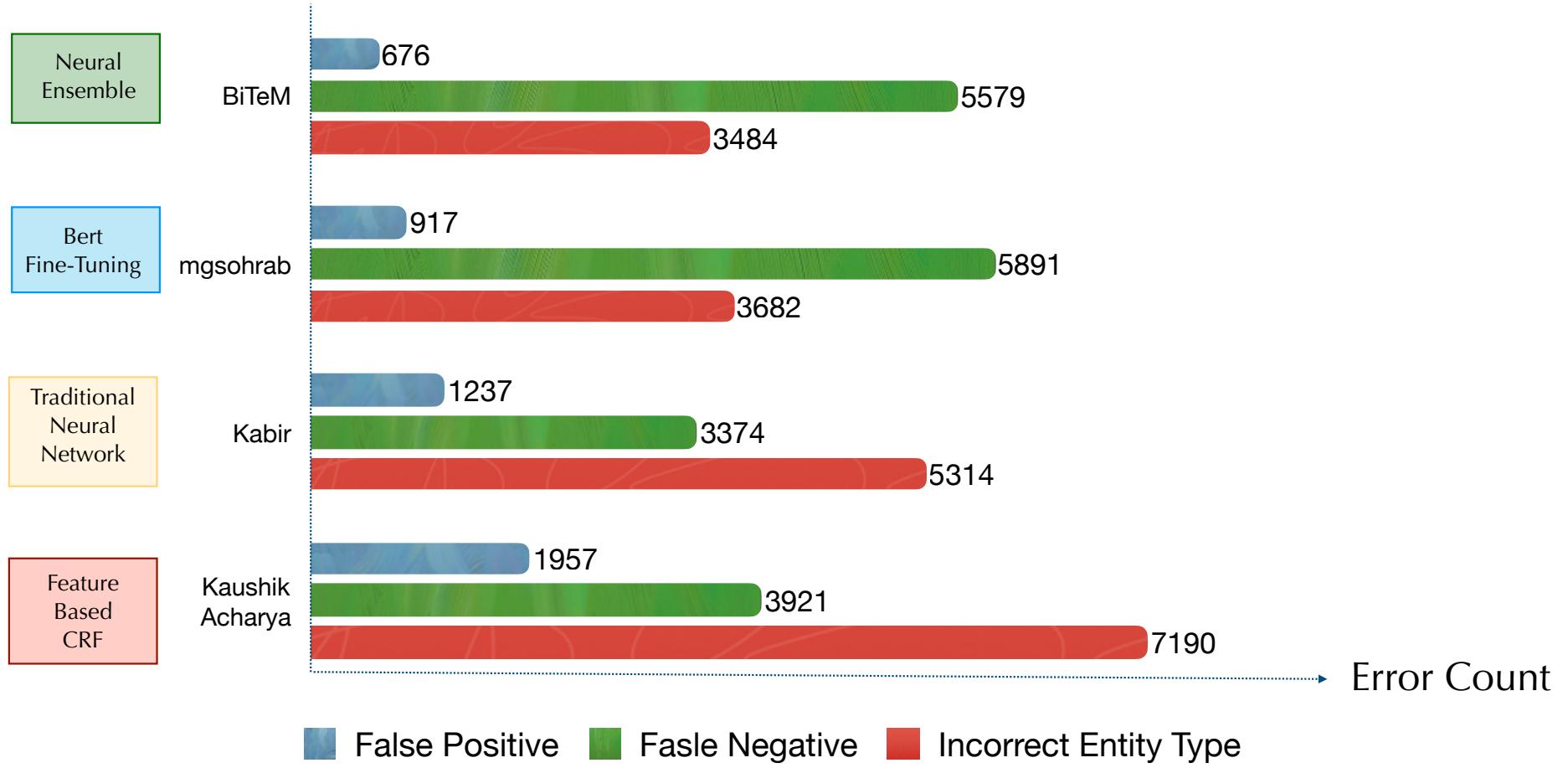
# NER Error Distribution



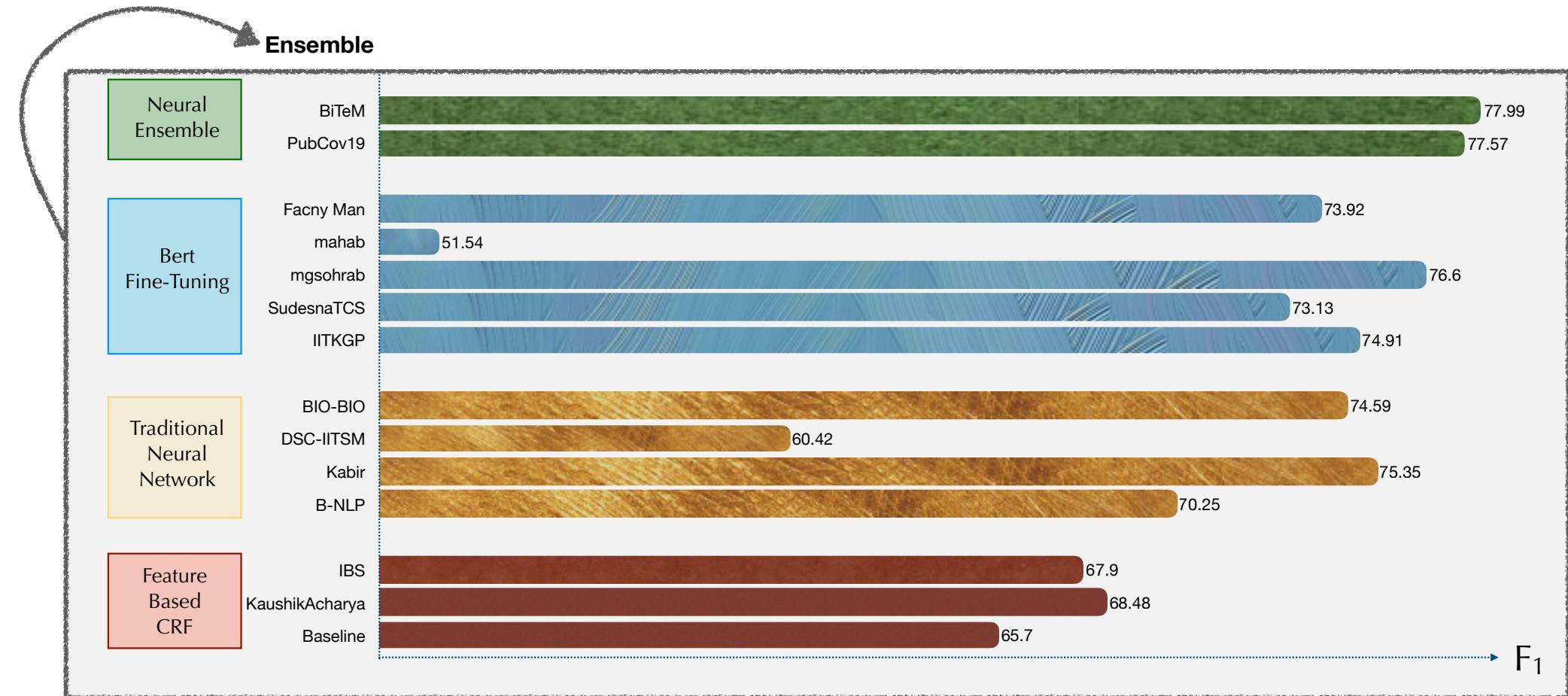
# NER Error Distribution



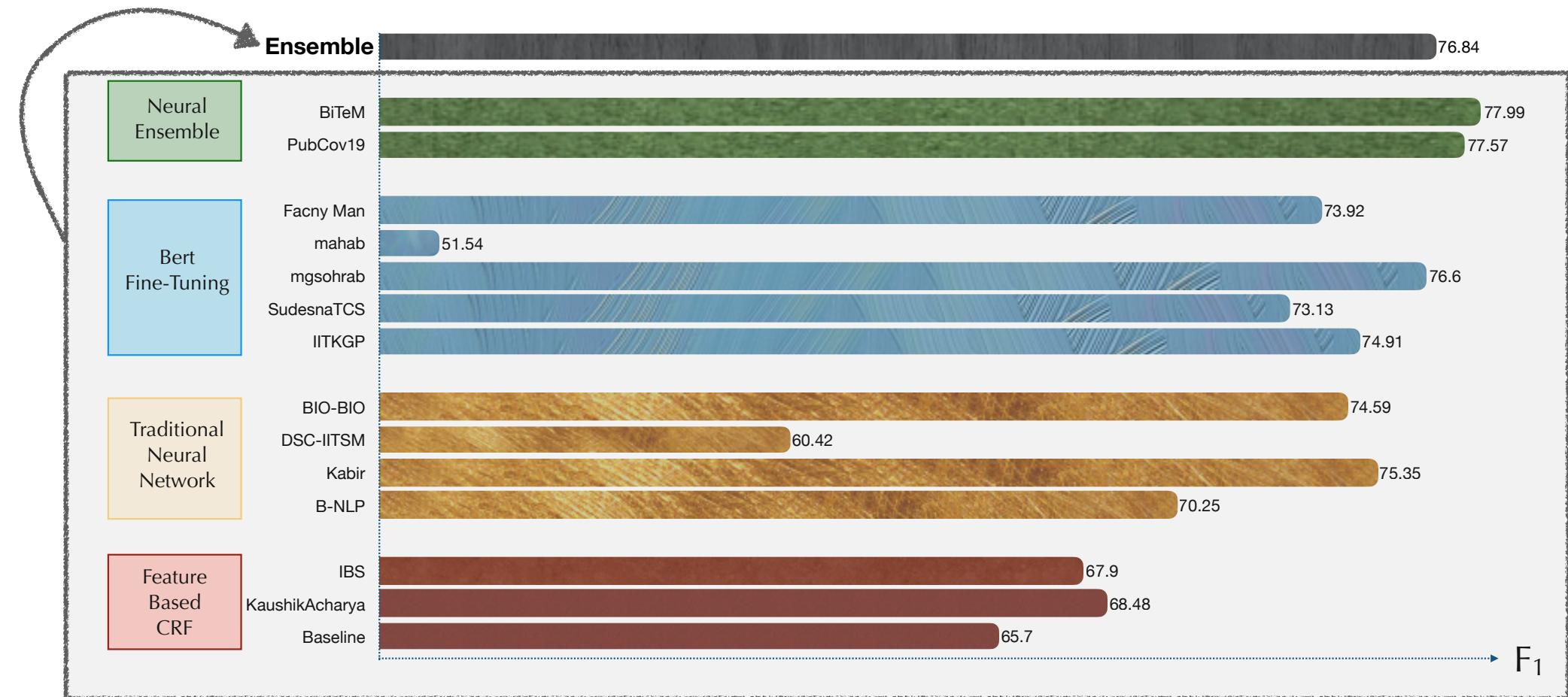
# NER Error Distribution



# NER Ensemble



# NER Ensemble



# Shared Task @WNUT-EMNLP-2020

## Sub-Task-1

 Mix the following components in an 0.2mL 8-strip tube;

Reagent      Amount      Location

Named Entity Extraction  
(NER)

## Sub-Task-2

    Mix the following components in an 0.2mL 8-strip tube;

Site      Reagent      Amount      Location

Relation Extraction  
(RE)



protocols.io

Archive timeline:  
2012 - 2020

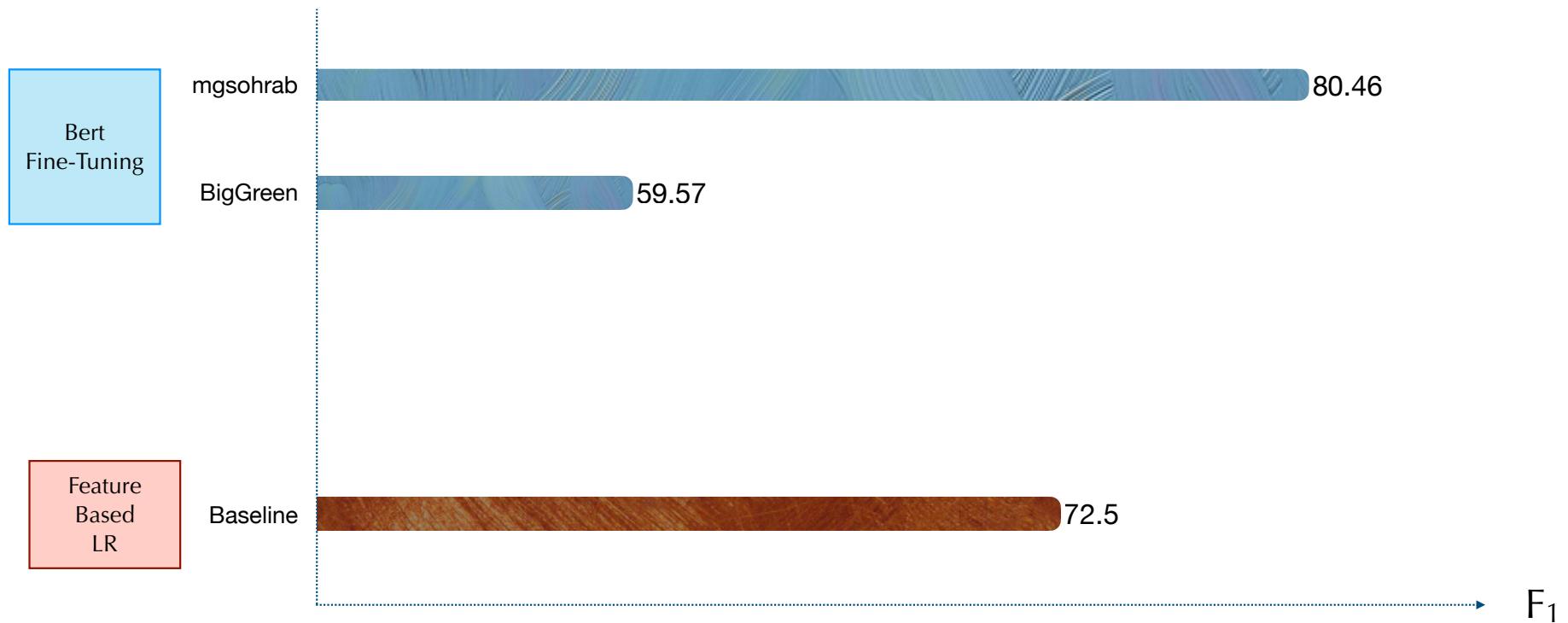
→ 726 protocols



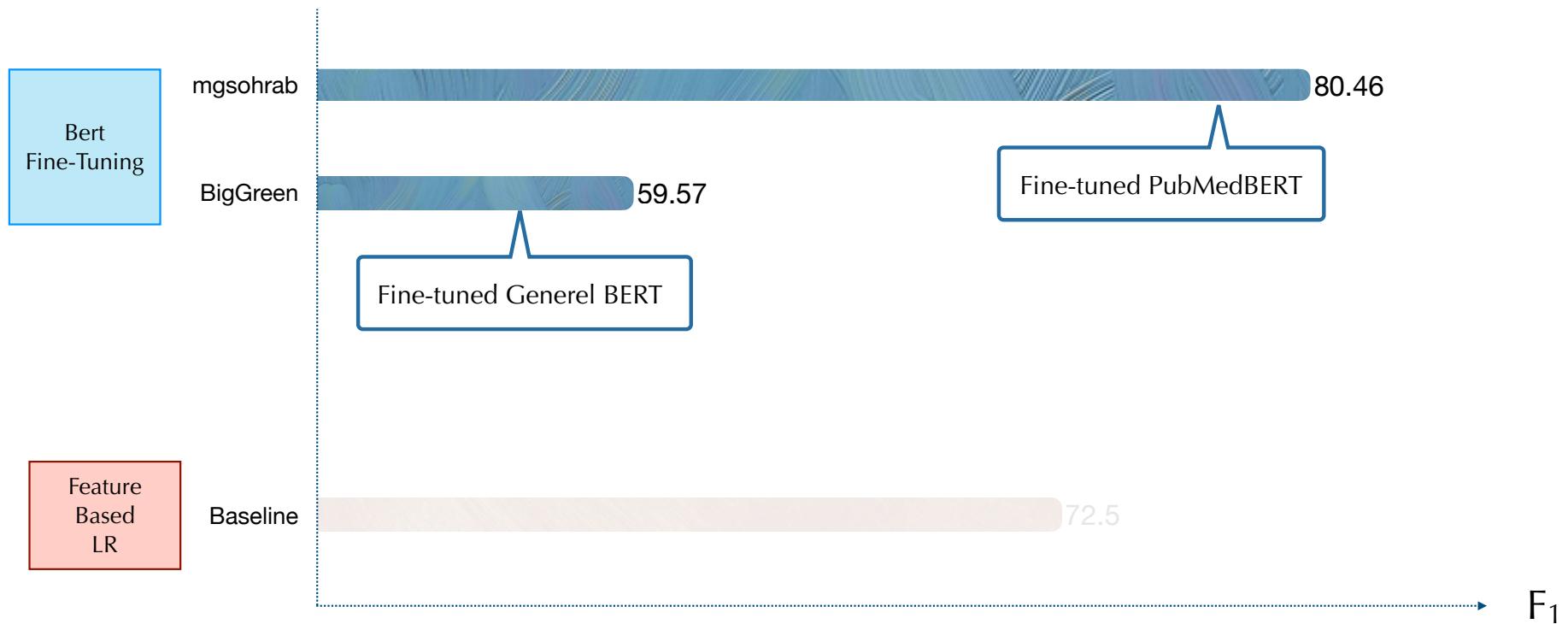
Train/Dev Data  
(14,096 Sentences)

Test Data  
(3,562 Sentences)

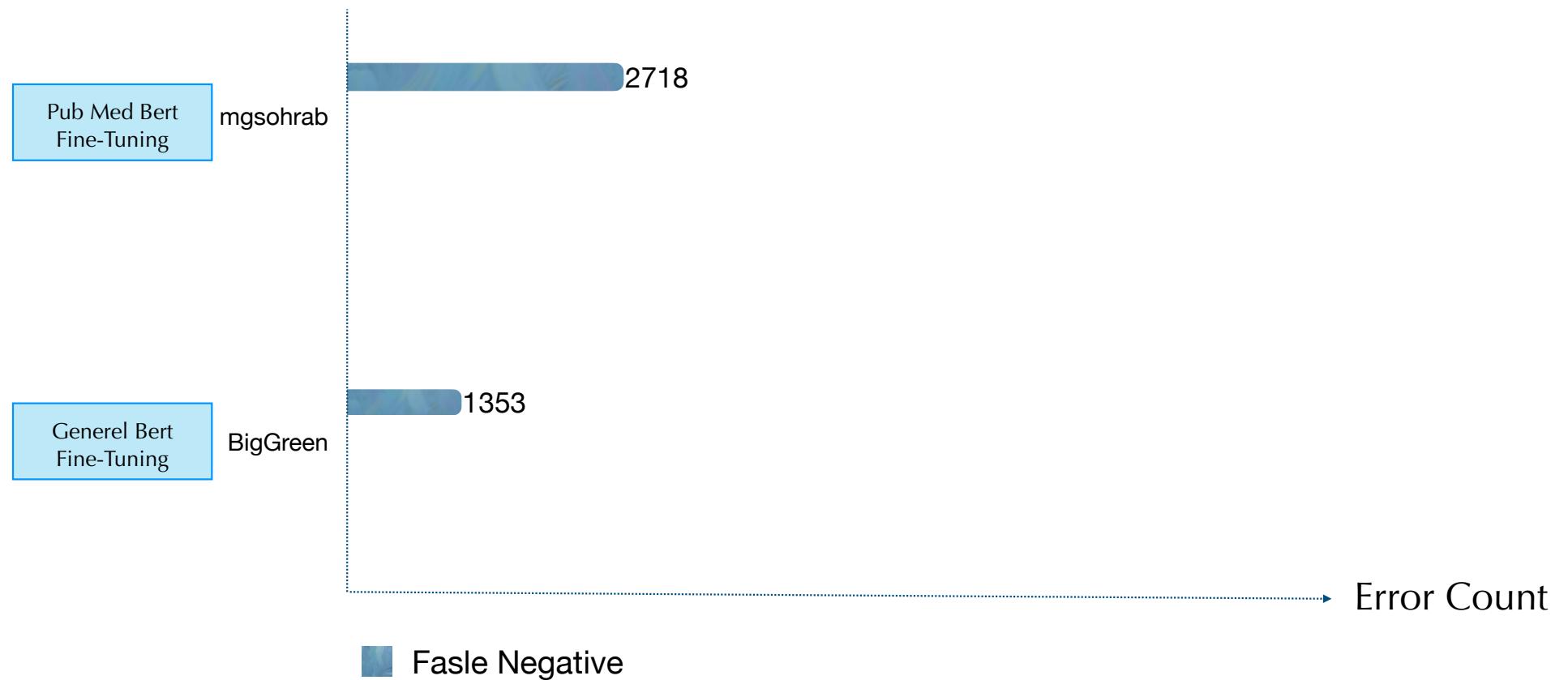
# RE Techniques



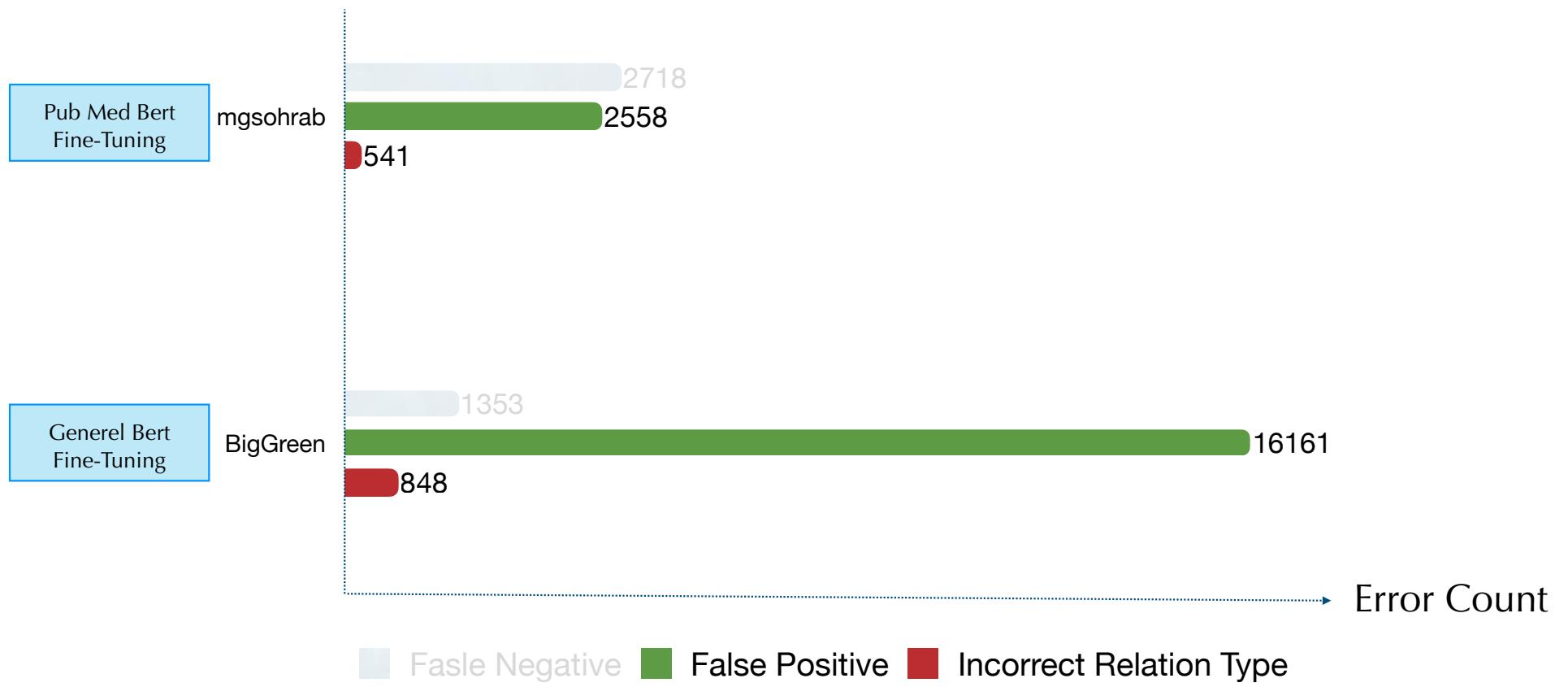
# RE Techniques



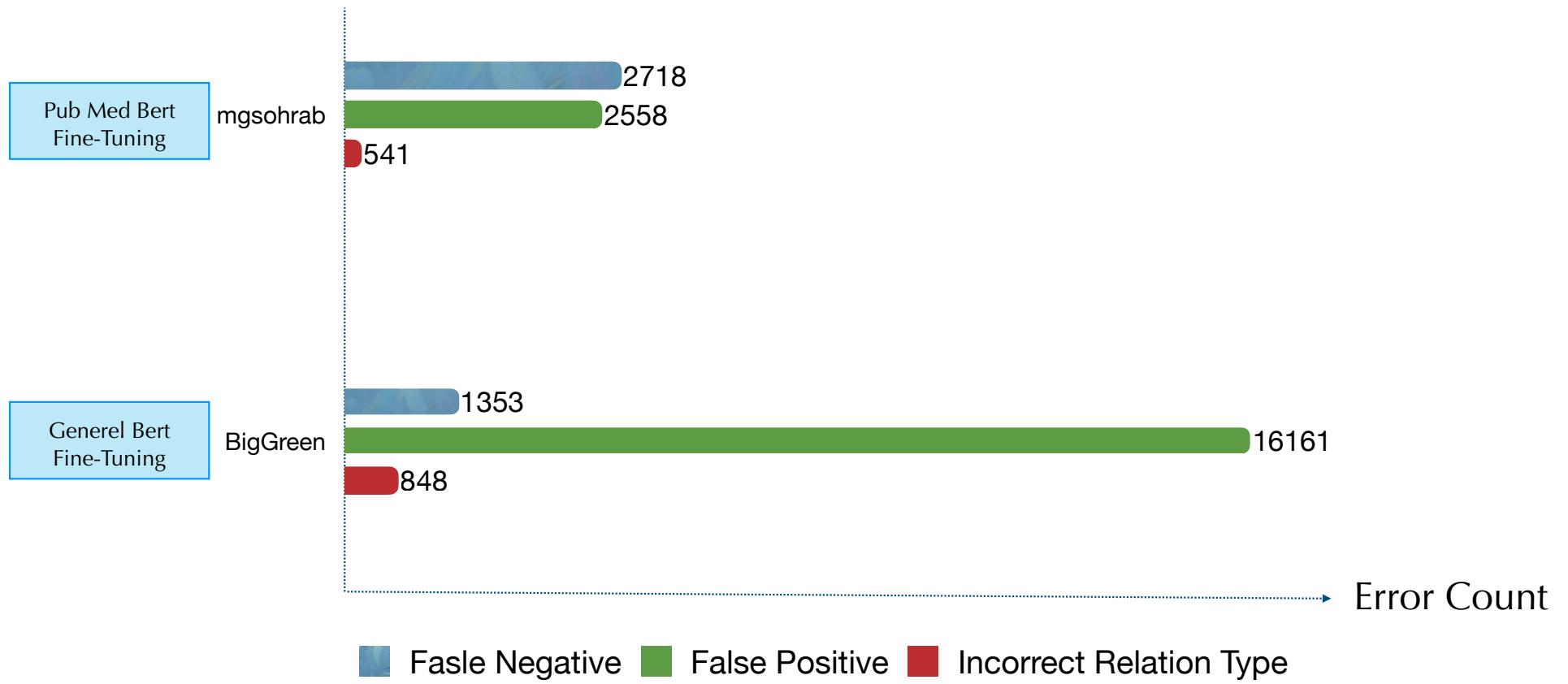
# RE Error Distribution



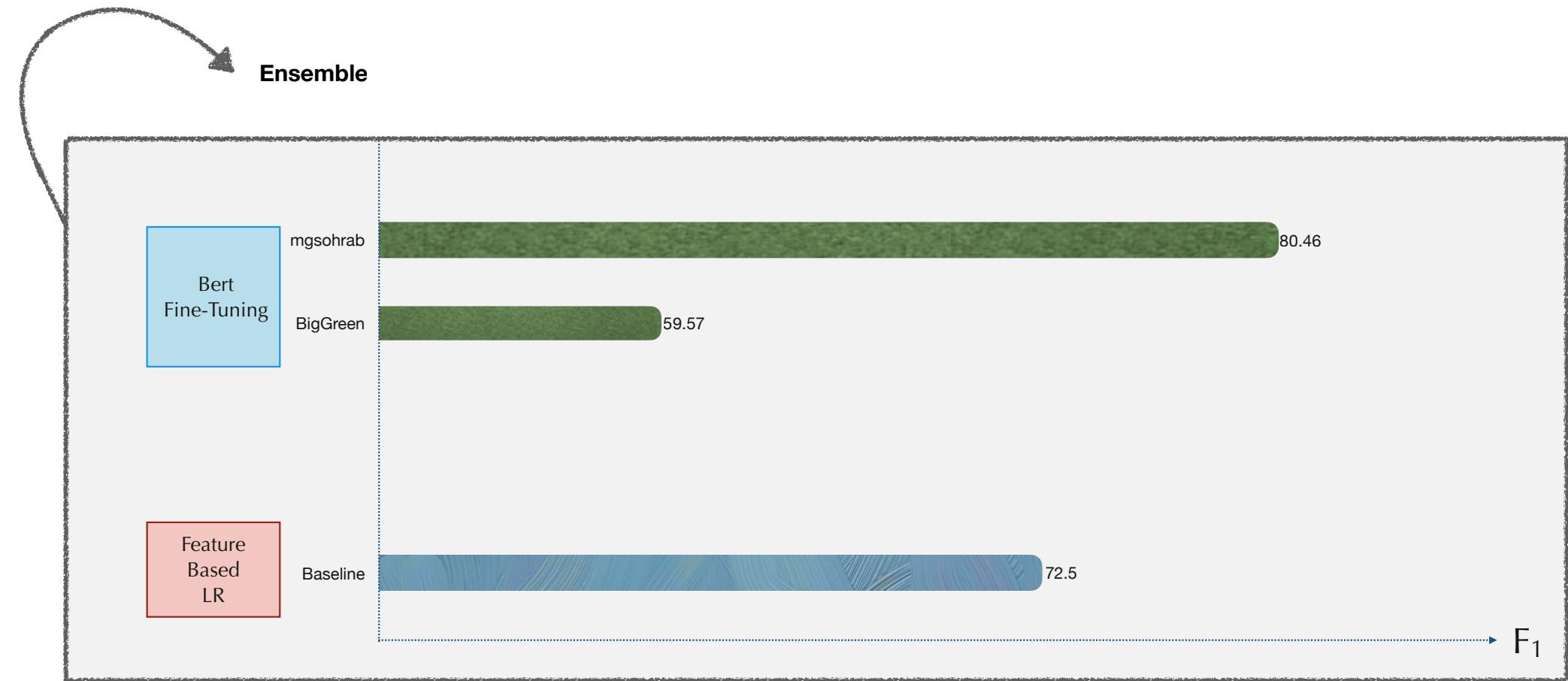
# RE Error Distribution



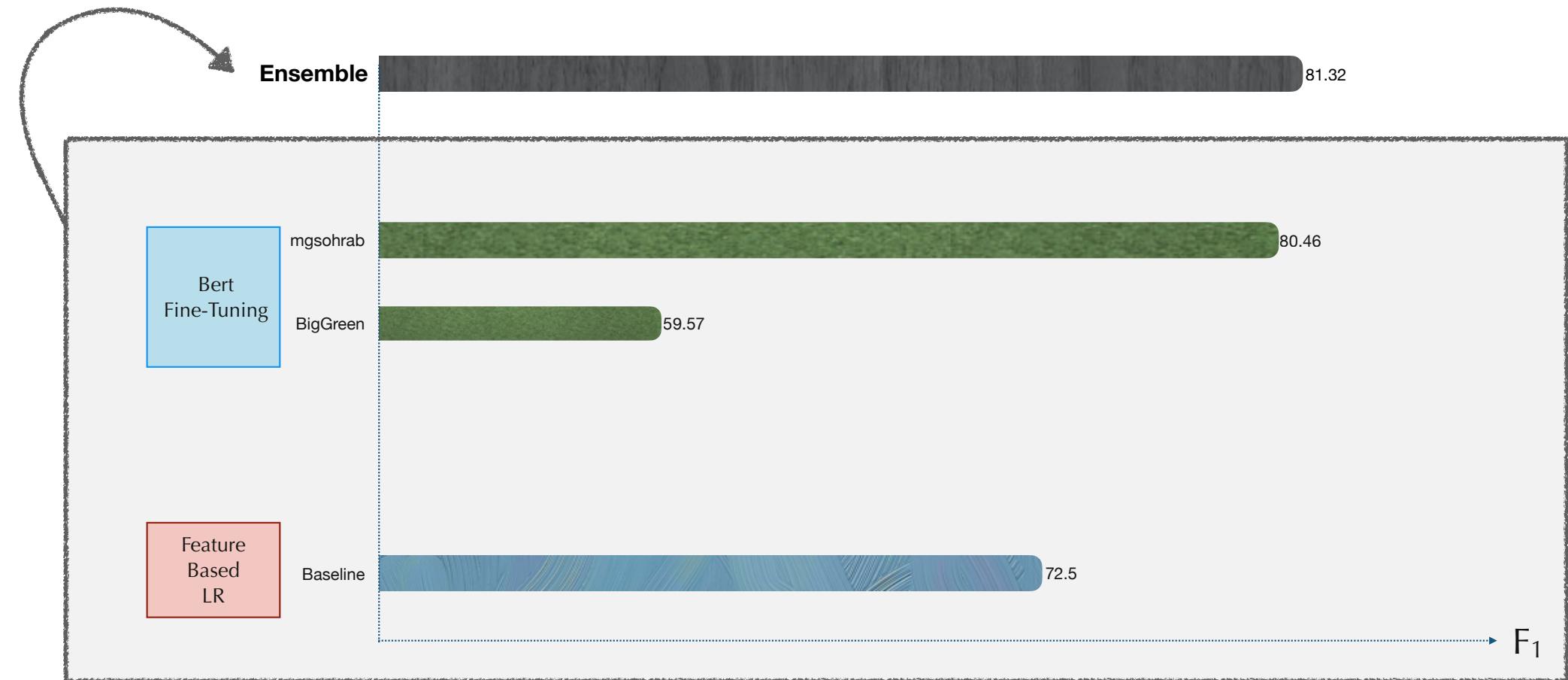
# RE Error Distribution



# RE Ensemble



# RE Ensemble



# Summary

---

Domain related BERT models boost performance for both the RE and NER task

---

# Summary

---

Domain related BERT models boost performance for both the RE and NER task

---

Ensemble of multiple BERT-models outperforms single fine tuned BERT for both the RE and NER task

---

# Summary

Poster Session  
@ET-12:35 PM  
@Room-E20



Domain related BERT models boost performance for both the RE and NER task

Ensemble of multiple BERT-models outperforms single fine tuned BERT for both the RE and NER task

Dataset available at: [https://github.com/jeniyat/WNUT\\_2020\\_NER](https://github.com/jeniyat/WNUT_2020_NER)

# Summary

Poster Session  
@ET-12:35 PM  
@Room-E20



Domain related BERT models boost performance for both the RE and NER task

Ensemble of multiple BERT-models outperforms single fine tuned BERT for both the RE and NER task

Dataset available at: [https://github.com/jeniyat/WNUT\\_2020\\_NER](https://github.com/jeniyat/WNUT_2020_NER)