MPP: Model Performance Predictor

**Sindhu Ghanta**, Sriram Subramanian, Lior Khermosh, Harshil Shah, Yakov Goldberg, Swaminathan Sundararaman, Drew Roselli, Nisha Talagala

*Work done at ParallelM*
Agenda

Monitoring production models

MPP for monitoring

Results

| Dataset | Primary Algorithm Accuracy | MPP predicted Accuracy | Absolute difference |
|---------|---------------------------|------------------------|---------------------|
| Samsung | 0.92                      | 0.92                   | 0                   |
| Yip     | 0.95                      | 0.95                   | 0                   |
| Censurs | 0.78                      | 0.63                   | 0.15                |
| Forest  | 0.65                      | 0.64                   | 0.01                |
| Letter  | 0.71                      | 0.6                    | 0.11                |
Monitoring Production Models

- Dataset (features, labels)
- Generate an ML model
- Performance metrics

Example:
- Dataset (Sensor records, activity)
- An activity detection model
- Accuracy = 98%
Assumption: The models training performance is indicative of its performance on production data
Monitoring Production Models

ML Engineer → Sandbox → Production → Operations

Deployment

Monitoring
MPP for monitoring

Features

Primary Model

Prediction

MPP Model

Primary Model’s Performance Prediction
MPP for monitoring

- Dataset (Sensor records, activity)
- An activity detection model
- Accuracy = 98%

Training example:
- Dataset (Sensor records, activity)
- An activity detection model
- Accuracy = 98%

Sensor Records

Primary Model

MPP Model

Activity

(Example: walking, running, sitting etc)

Binary value

1: Primary model made an accurate prediction

0: Primary model prediction is incorrect
MPP for monitoring

ML Engineer

Sandbox

Deployment

Production

Operations

Monitoring

MPP Model
How to create a MPP model

1. Training Dataset
2. Training
3. Primary Model
4. Test Data
5. Predictions
How to create a MPP model
## Results

| Dataset | Primary Algorithm Accuracy | MPP predicted Accuracy | Absolute difference |
|---------|-----------------------------|------------------------|---------------------|
| Samsung | 0.92                        | 0.92                   | 0                   |
| Yelp    | 0.95                        | 0.95                   | 0                   |
| Census  | 0.78                        | 0.63                   | 0.15                |
| Forest  | 0.65                        | 0.64                   | 0.01                |
| Letter  | 0.71                        | 0.6                    | 0.11                |

All datasets are publicly available in UCI repository
ML model to monitor an ML model?

Who monitors MPP’s performance?
Questions ?

Monitoring production models

MPP for monitoring

Results