Genome builder

Tokenize

- Choice 1 - name? - 0%
- Choice 2 - name? - 0%
- Choice 3 - name? - 0%

Pre-training

- Choice 1 - name? - 0%
- Choice 2 - name? - 0%
- Choice 3 - name? - 0%
- Choice 4 - name? - 0%
- Choice 5 - name? - 0%

1.0. Collect data 50%
1.1. Clean data 100%
1.2. Tokenize 30%
1.3. Pre-training 0%

95% Accuracy on 5 main categories from Arxiv

2.1. New class detection 0%
2.2. Force cluster 0%
2.3. ....
2.4. ....

Genome Builder is created

Collect real data

- Select providers - 0%
- Scrappers - 0%
- Maintain databses - 0%

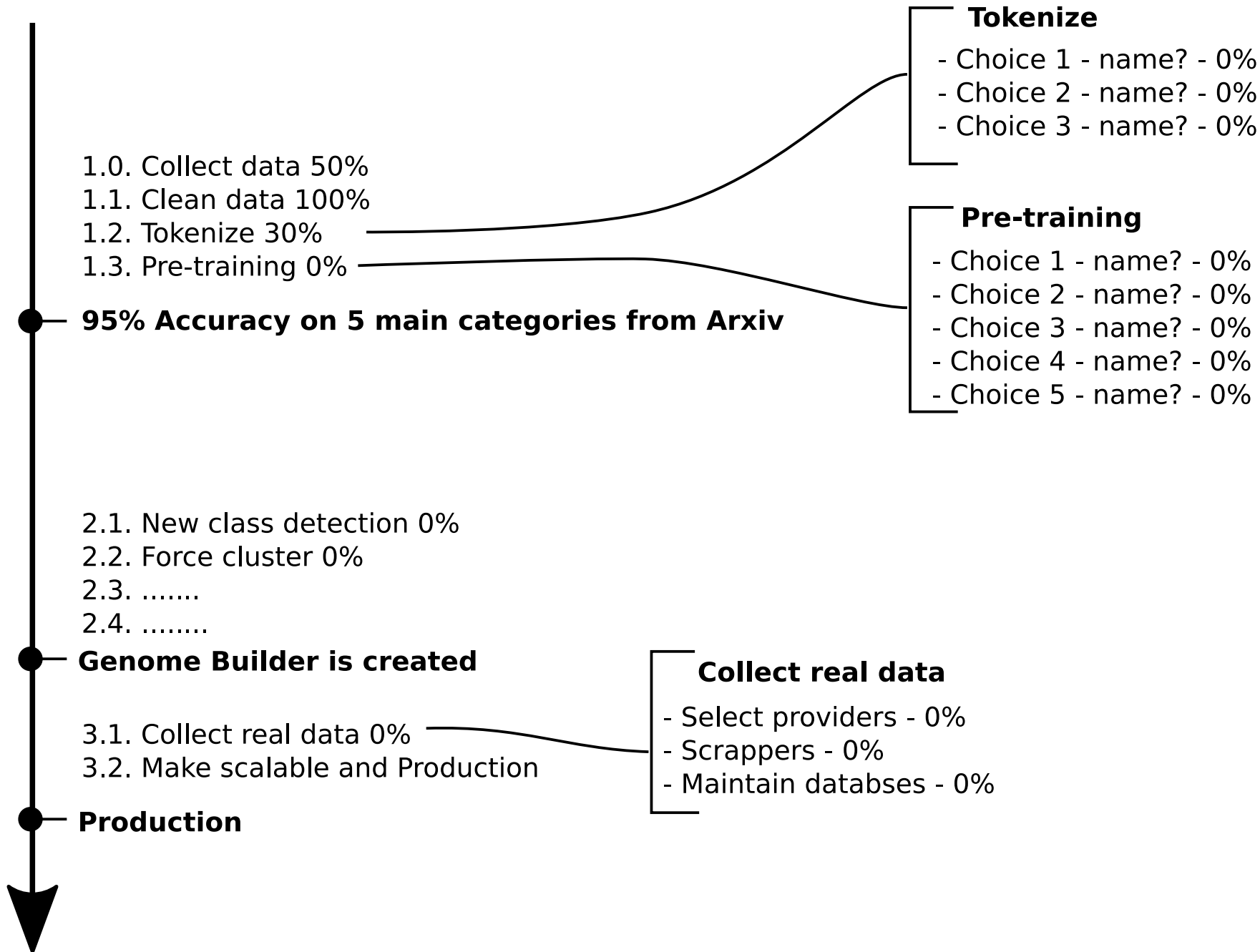3.1. Collect real data 0%
3.2. Make scalable and Production

Production

Knowledge Graph and Sentiment analysis Prototype

1.0. Find data for prototype - 0%
1.1. Clean data - 0%
1.2. Named Entitties Recognition (NER) - 0%
1.3. Select Entity for test - 0%
1.4 Apply Attention method to get Entity context - 0%
1.5 Calculate sentiments for specific Entity - 0%

Build Sentiment tool prototype

2.1. Aggregate mentions about Entity - 0%
2.2 Take all Entitites from NER and build KnowledgeGraph - 0%
2.3 Aggregate sentiments and mentions to all Entities in Knowledge Graph - 0%

Build Knowledge Grpah with all information

Find data for prototype

- Fast scrapping small data - 0%
- Find data exapmles - 0%

Prediction Module

1.0. Collect data (from Sesamm) - 100%
1.1. Cluster time-series (K-means, etc.)- 0%
1.2. Test prediction models - 0%
Build Predictive module

Test predicition models

- ARIMA/SARIMA - 0%
- LSTM - 0%

MODEL Test & Select

Group(n1,n2,n3)

GARCH

LSTM

Exponential smoothing

ARIMA, SARIMA

Nested Cross-Validation

Train | Test

Test

Test

Test

Test

Outer Loop
1. Train each split with optimal parameters
2. Average each split's test error

Training Subset | Validation

Inner Loop
Tune hyperparameters