

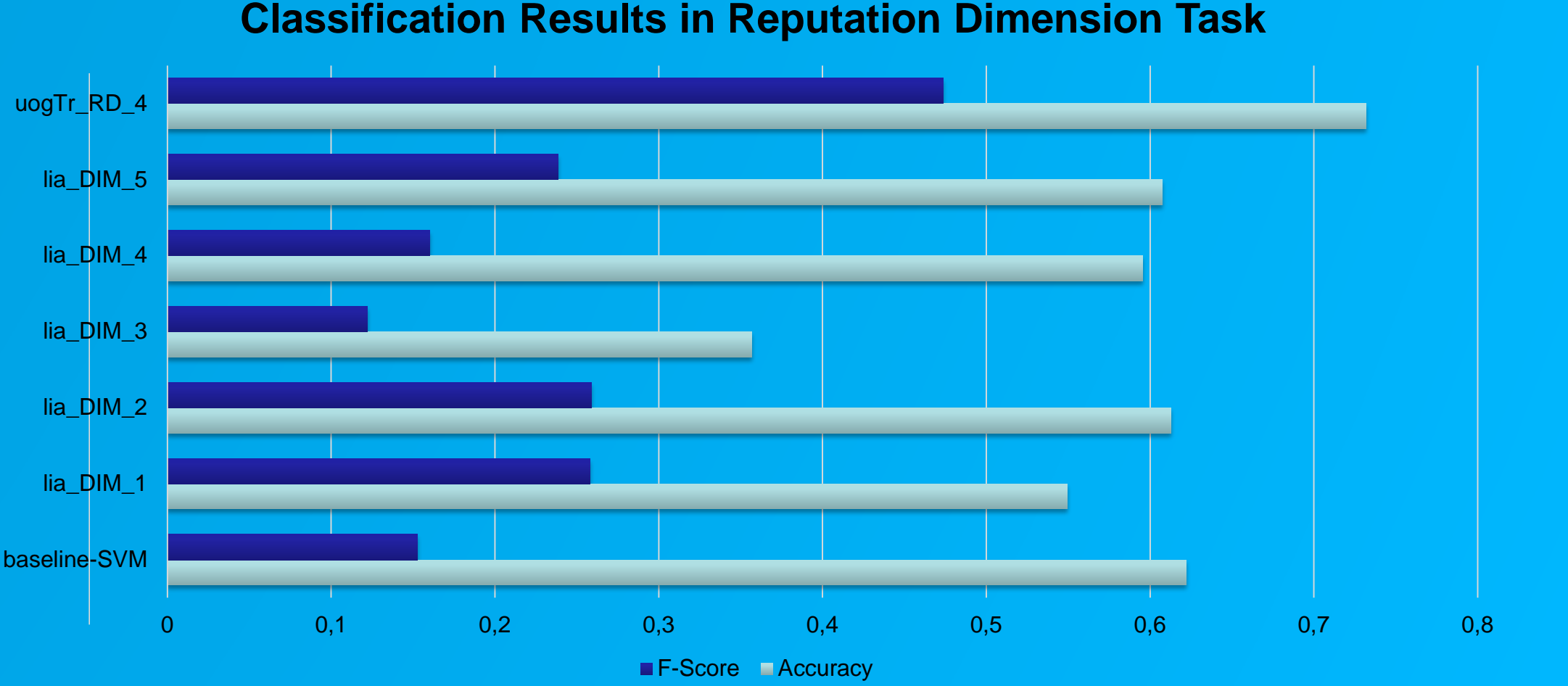
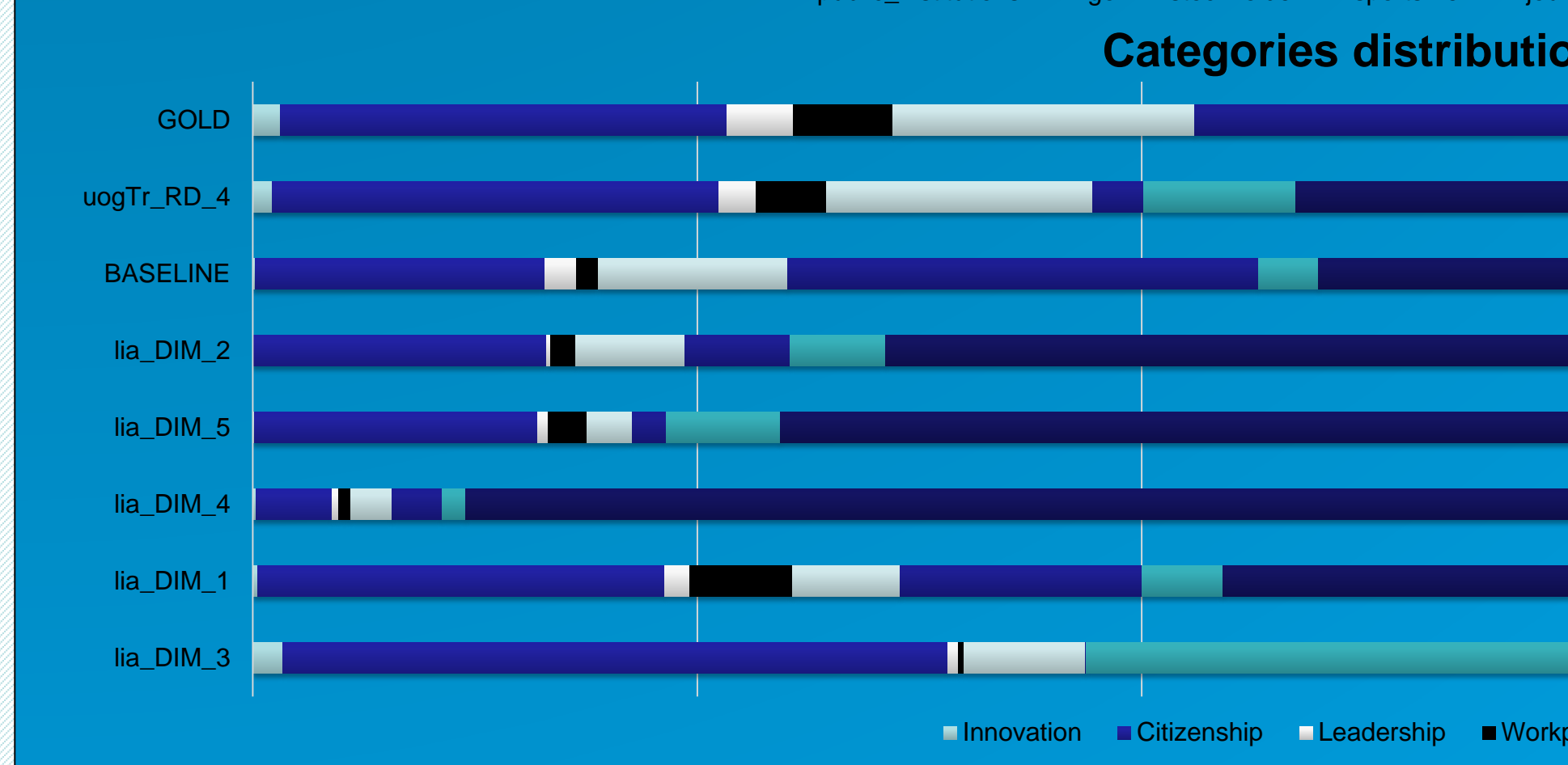
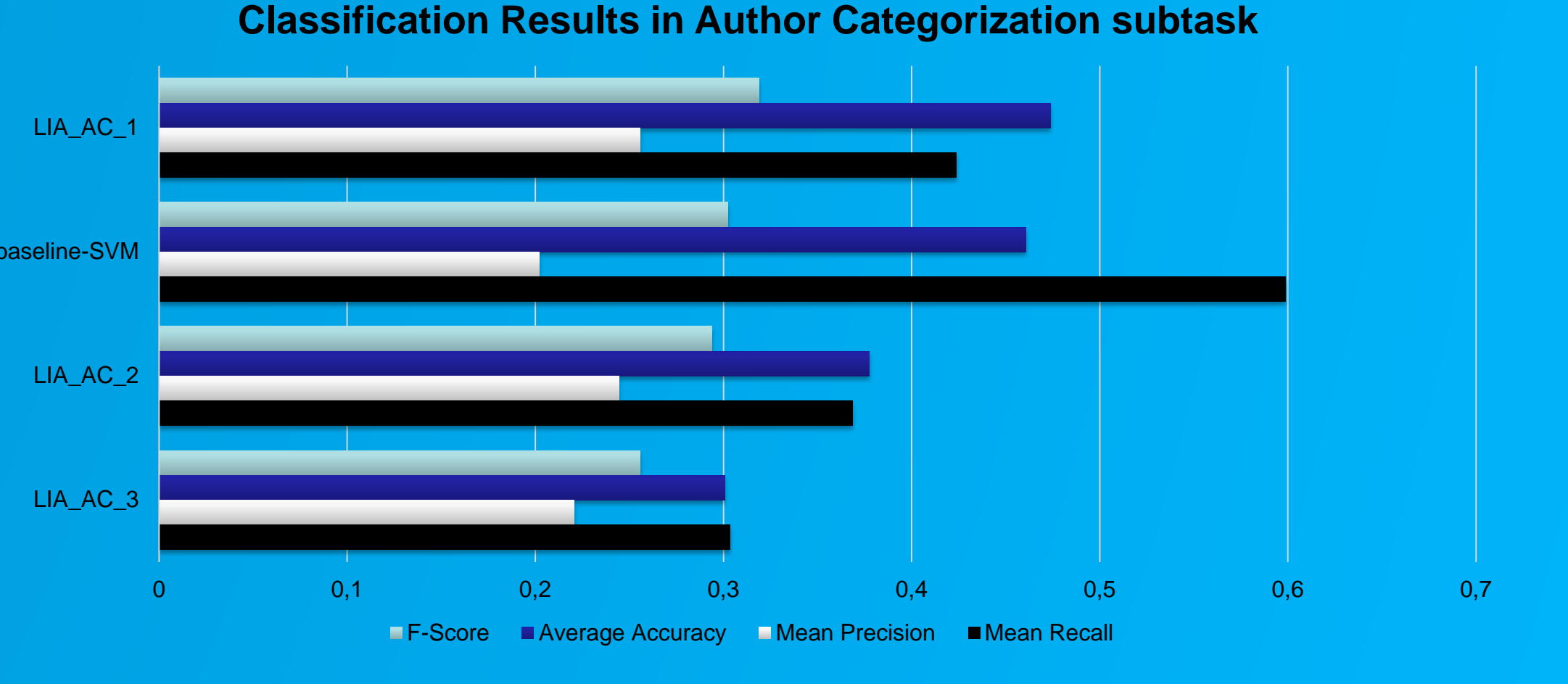
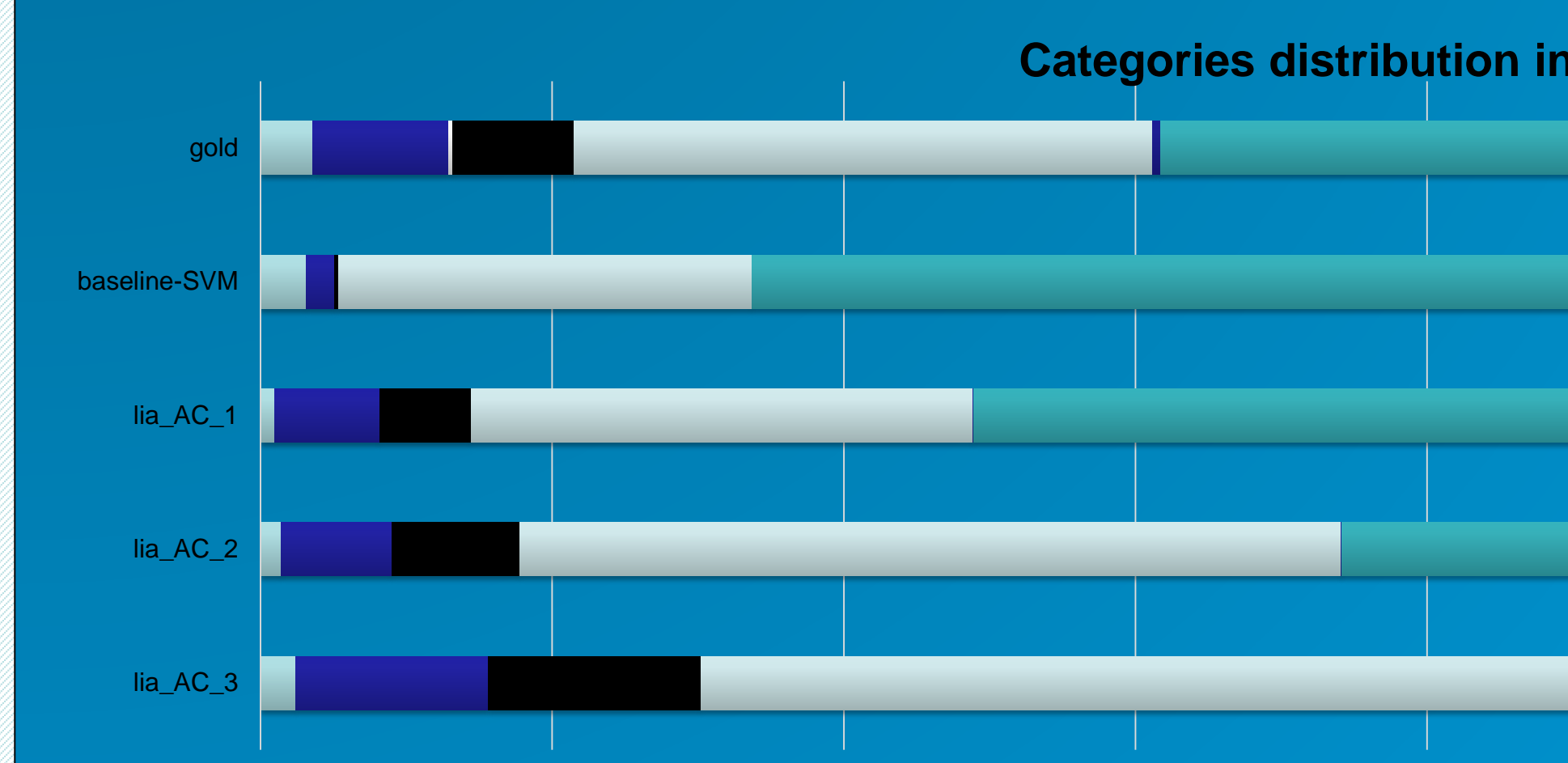
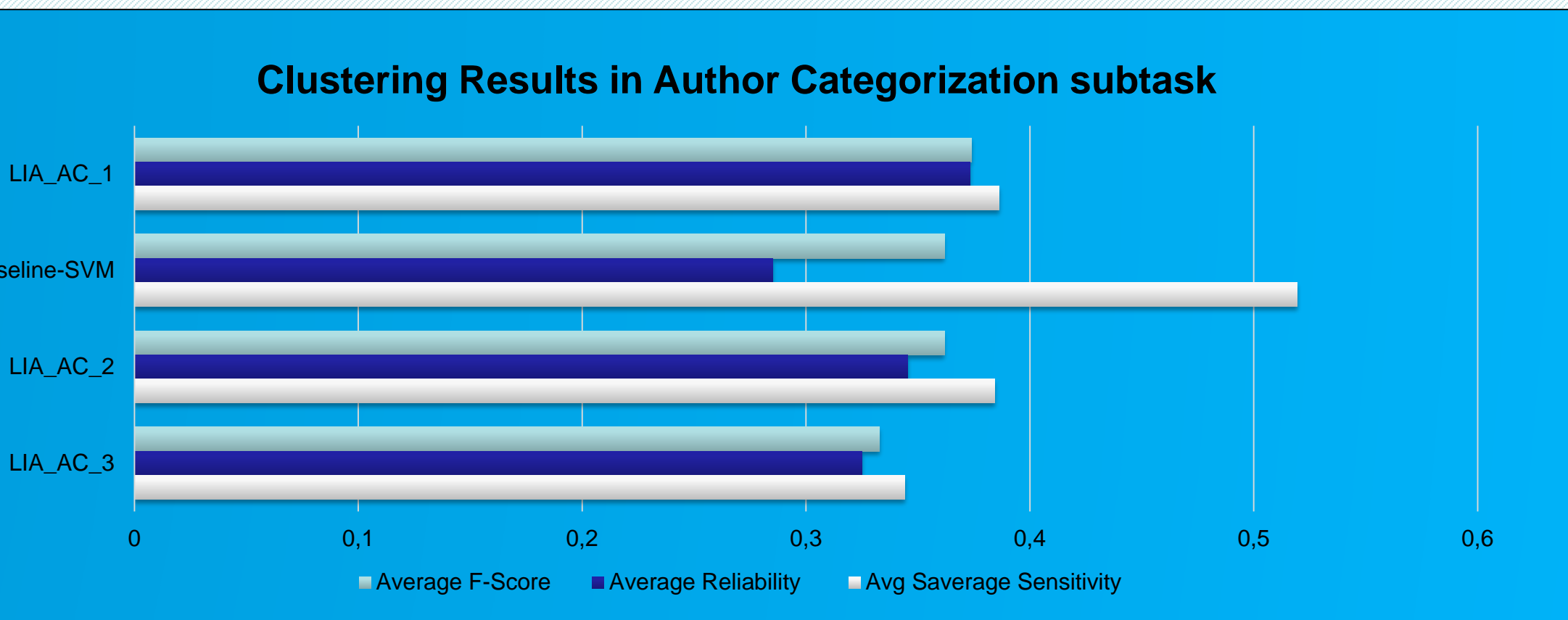
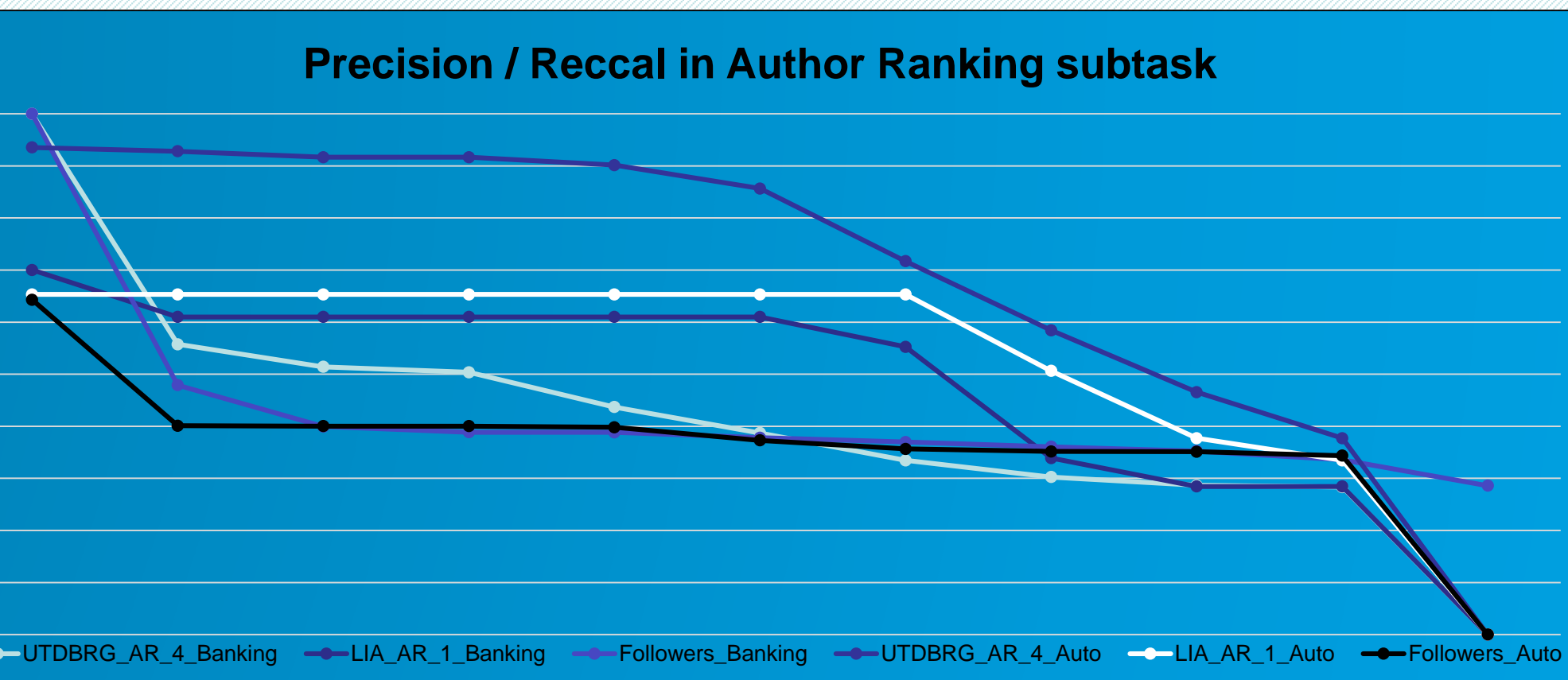
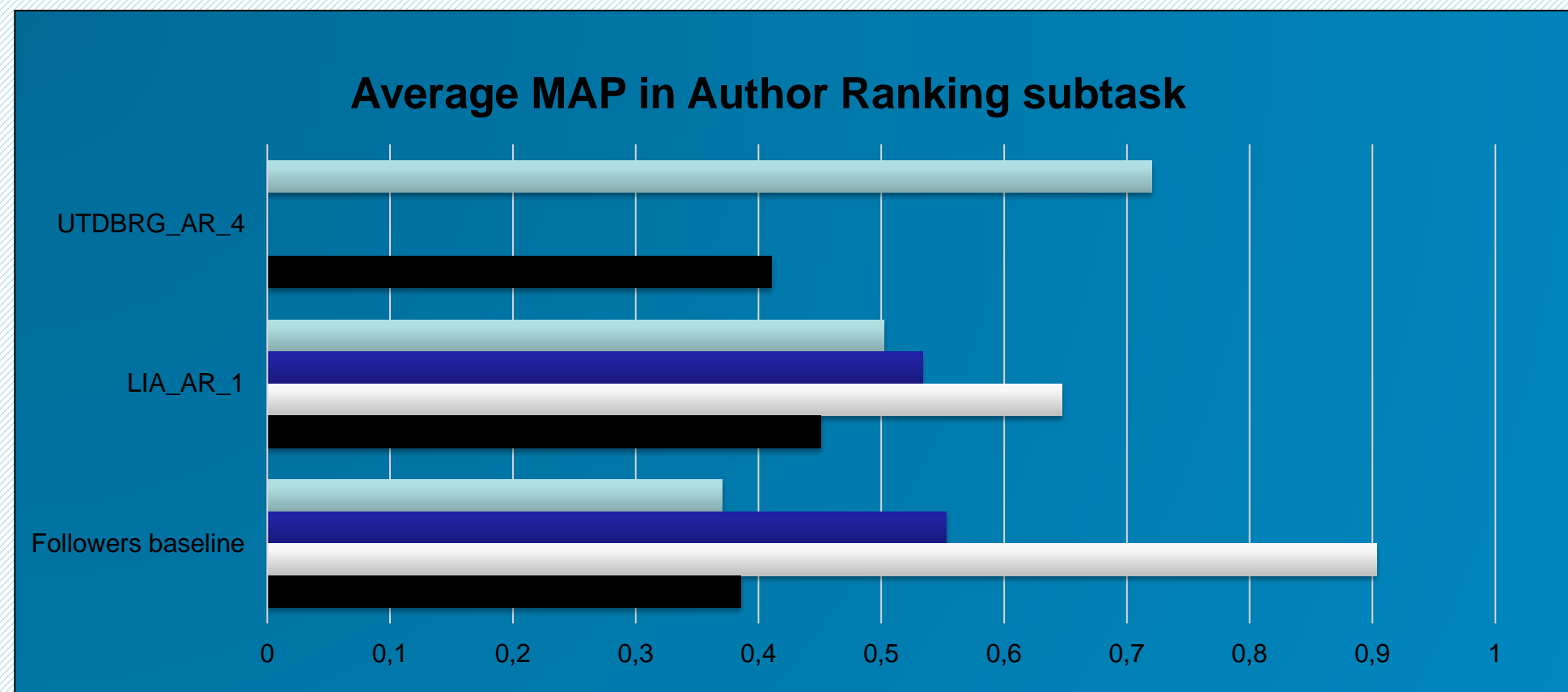
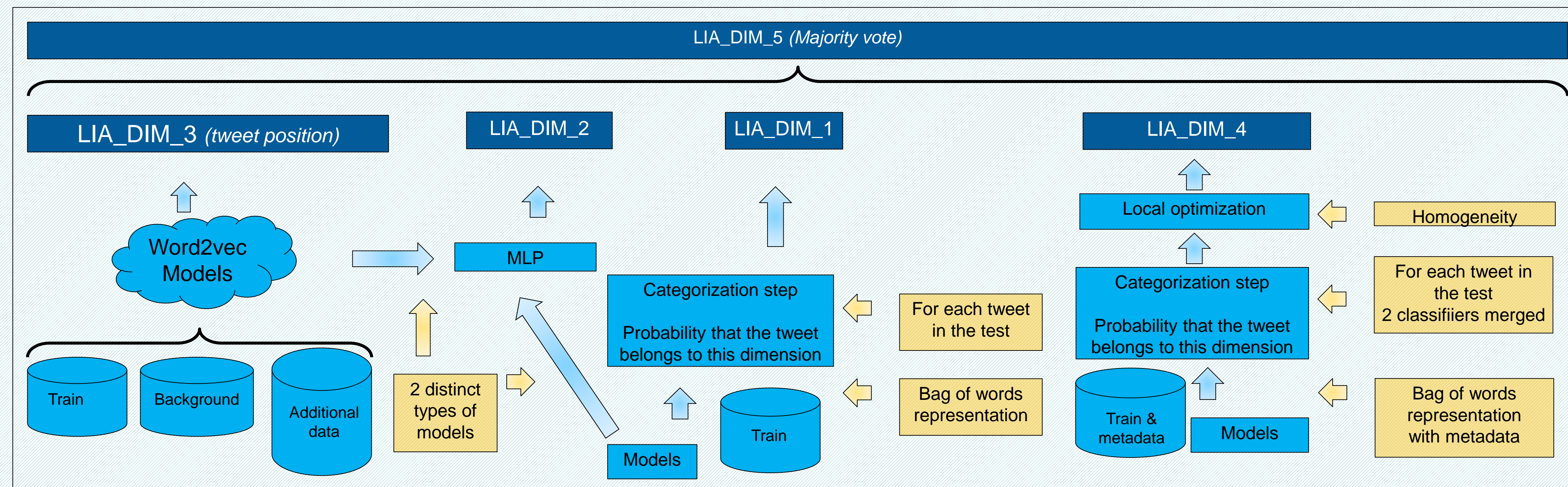
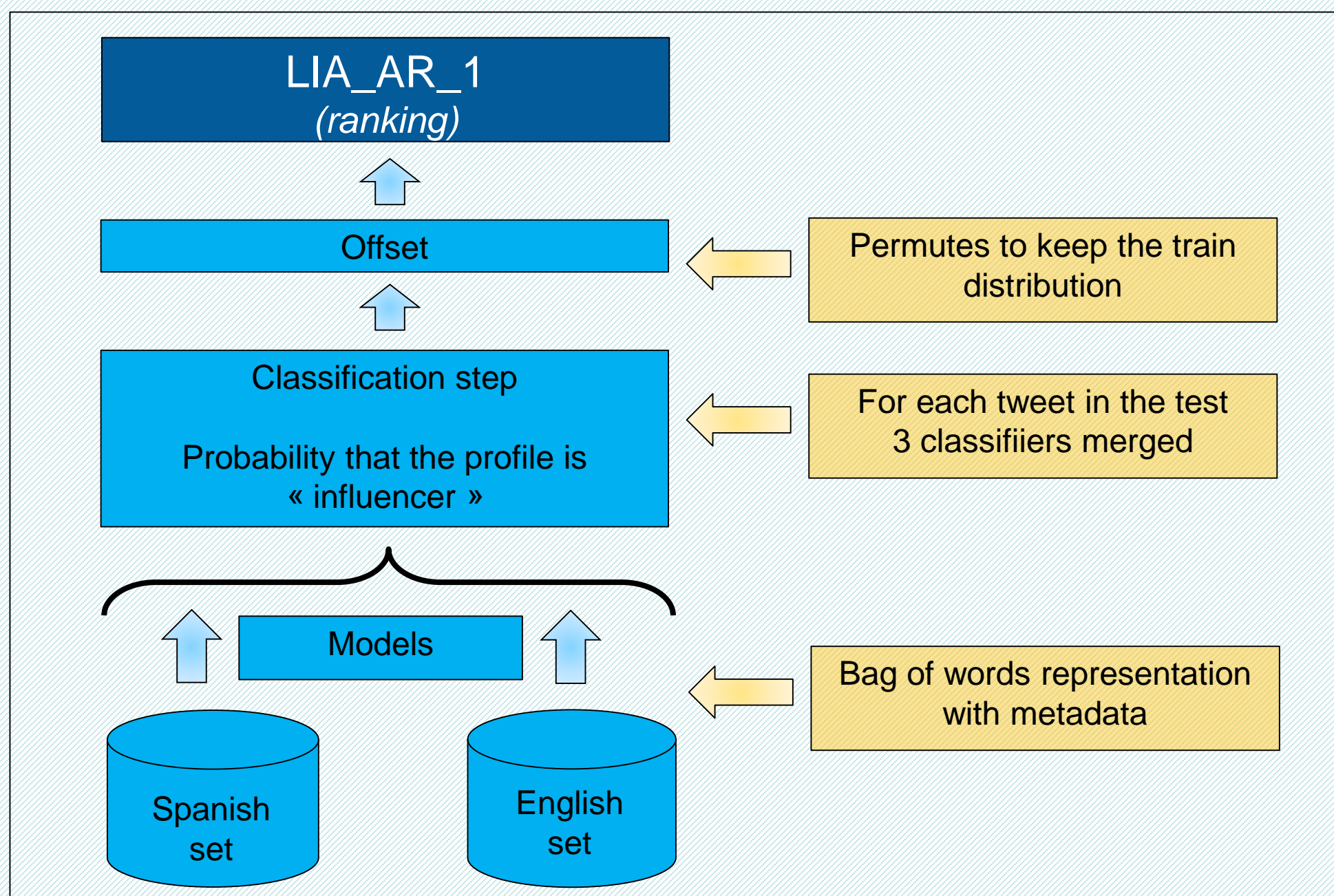
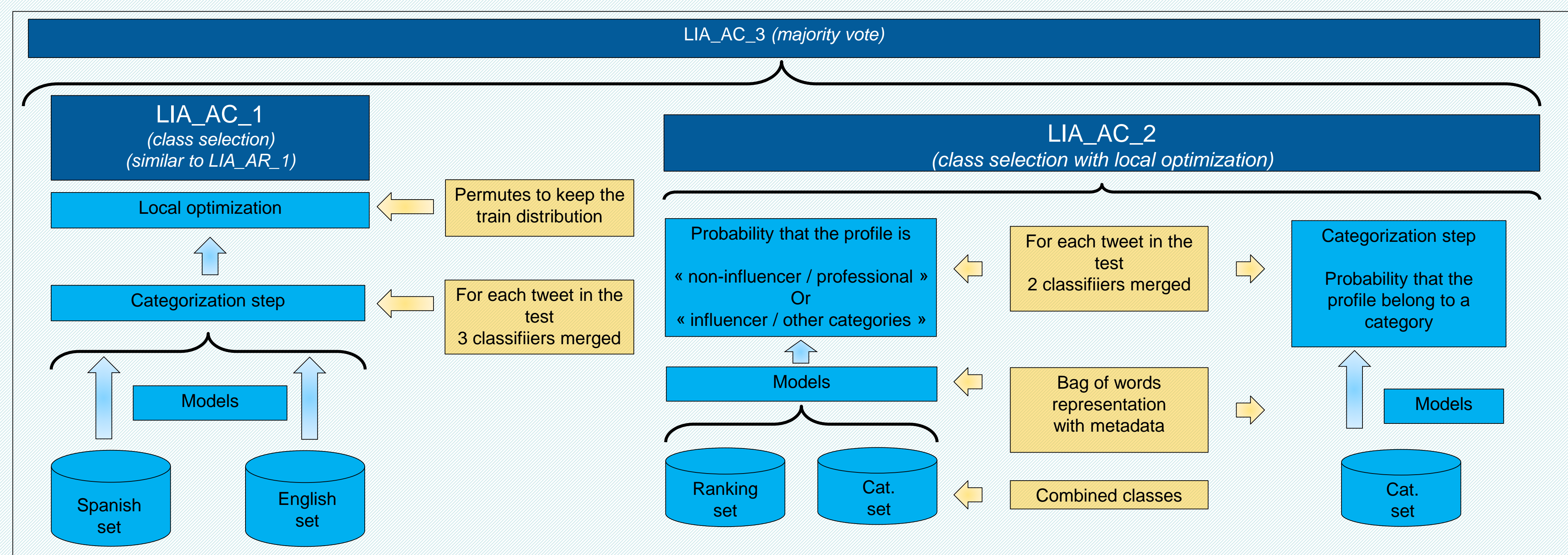
Motivations (1) try a large variety of speech and NLP approaches to tackle the reputation issue (2) use local outputs optimization

Submitted runs for each subtasks

1 run in Author Ranking
 lia_ar_1 : Merge of HMM, Poisson and Cosine (per lang specific models)

3 runs in Author Categoriization
 lia_ac_1 : Merge of HMM, Poisson and Cosine (per lang specific models)
 lia_ac_2 : Merge of HMM and Cosine (global models)
 lia_ac_3 : Merge of lia_ac_1 and lia_ac_2

5 runs in Reputation dimensions
 lia_dim_1 : Conditional random field
 lia_dim_2 : Multilayer Perceptron
 lia_dim_3 : Naive Word2vec
 lia_dim_4 : Merge of HMM and Cosine (global models)
 lia_dim_5 : Merge of lia_dim_1 to lia_dim_4



Conclusive analyses

Although our systems performed well in Author Profiling subtasks we may have missed something in the reputation dimensions task. By over-predicting the main class our systems failed to handle the main classification issue of the task if we consider that the main goal was to find the small classes.

Our author categorization runs are close-call with the baseline at the same level in both classification and clustering evaluation. There is still a room for improvements but the performance level yield by our systems means that researchers are on a good way to tackle the problem.

The classification system used in the author ranking subtask performed competitively with regards to the other participants and the baseline whereas it did not produced a real ranking output.

Combinations (LIA_AC_3 and LIA_DIM_5) are still lower than each system taken alone while we observed that each one brings good information at the entity level.

References

Amigo & al : Overview of RepLab 2013

Bahl & al : Obtaining candidate words by polling in a large vocabulary speech recognition system (1988)

Lafferty & al : Conditional random fields: Probabilistic models for segmenting and labeling sequence data (2001)

Mikolov & al : Efficient Estimation of Word Representations in Vector Space (2013)

Torres-Moreno & al : Opinion detection as a topic classification problem (2013)

Wang & Li : Automatic Text Classification Based on Hidden Markov Model and Support Vector Machine (2013)

Perspectives (1) Automatic summarization of tweets clusters and profiles (2) learn to rank with background data