

Paper presentation

PipelineProfiler: A Visual Analytics Tool for the Exploration of AutoML Pipelines

- Evaluation

GROUP

- Fjodor Ševtšenko
- Lisanna Lehes
- Marilin Moor
- Dmitri Rozgonjuk

Explainable Automated Machine Learning, LTAT.02.023

Evaluation

Case studies data

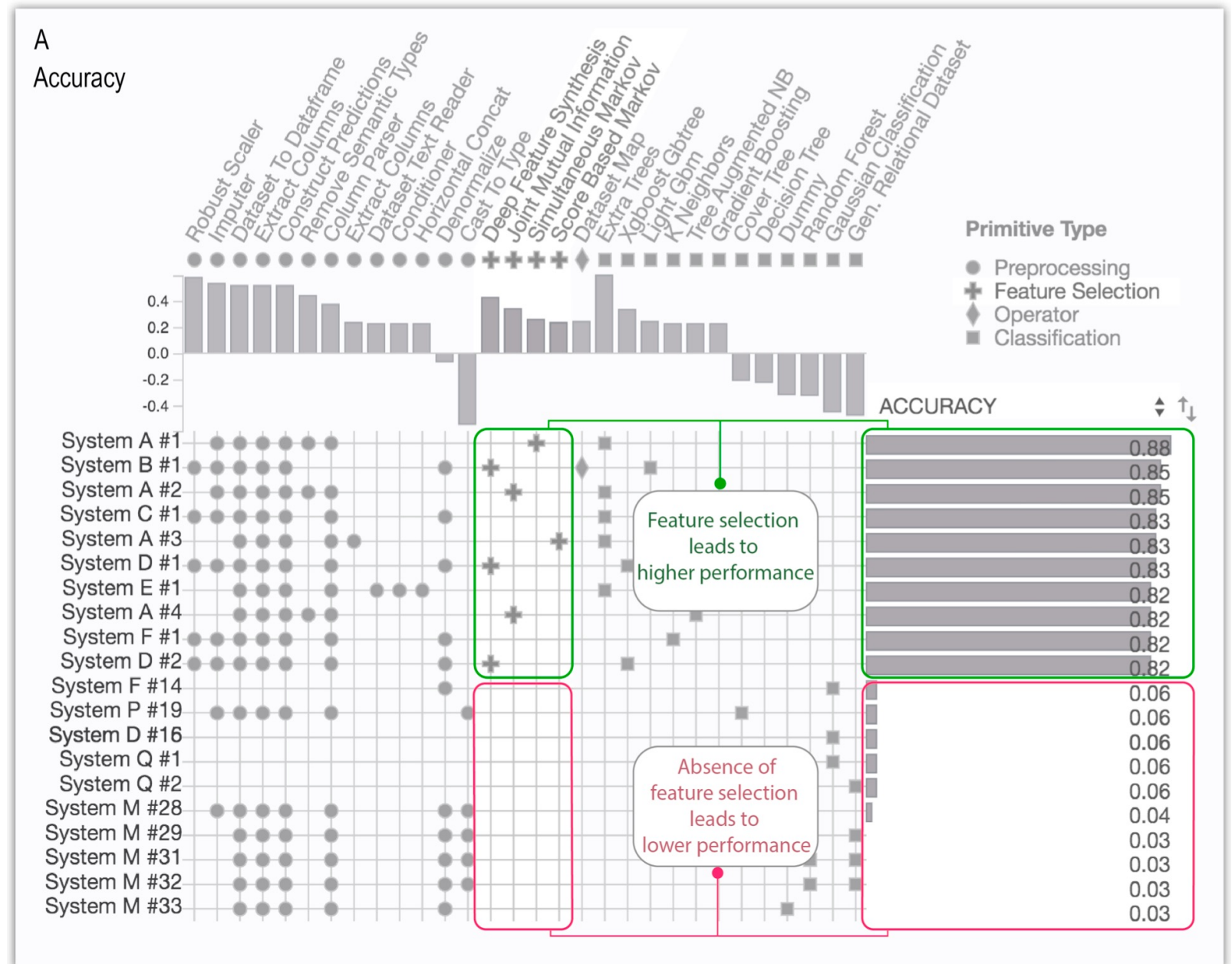
- D3M Program Summer 2019 Evaluation
- Created collection of data contains **10131 pipelines**
- 20 AutoML systems were run to solve various ML tasks:
 - Classification; Regression; Graph matching; Link prediction; Object detection
- Over **40 datasets** that cover multiple data types:
 - Tabular; Image; Audio; Time-series; Graph
- Each AutoML system was executed for one hour and derived zero or more pipelines for each dataset.

Case Study 1: Improving an AutoML System

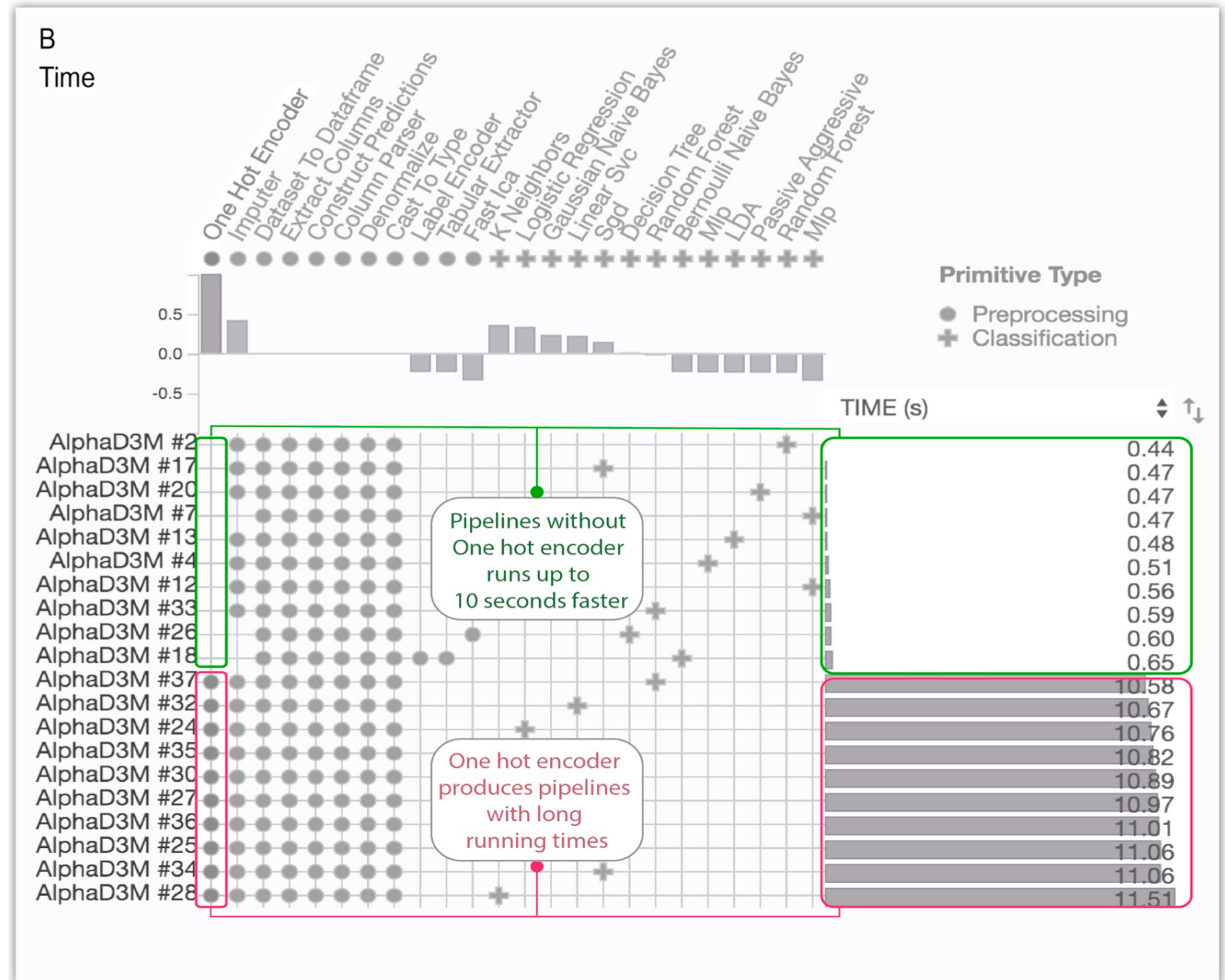
- The AlphaD3M developer started his exploration using a problem for which AlphaD3M¹ had a poor performance:
 - A **multi-class classification task** using the libras move dataset from the OpenML database.
- For this task, in the ranking of all pipelines produced by D3M AutoML systems, the best pipeline produced by AlphaD3M was ranked 18th with an accuracy score of **0.79**.
- After the analysis, the developer modified the AlphaD3M system's handling of **feature selection**, the **one-hot encoding primitive**, and the **prioritization of primitives**.
 - The new version of AlphaD3M now leads the ranking for the multi-class classification task in the libras move dataset with accuracy of **0.88**.
 - With respect to execution time, the current average time to evaluate each pipeline for this problem is less than **1 second**, while previously it took 10 seconds.

¹ AlphaD3M is an AutoML system based on reinforcement learning that uses a grammar (set of primitives' patterns) to reduce the search space of pipelines.

Pipelines Matrix (4A)



Pipelines Matrix (4B)

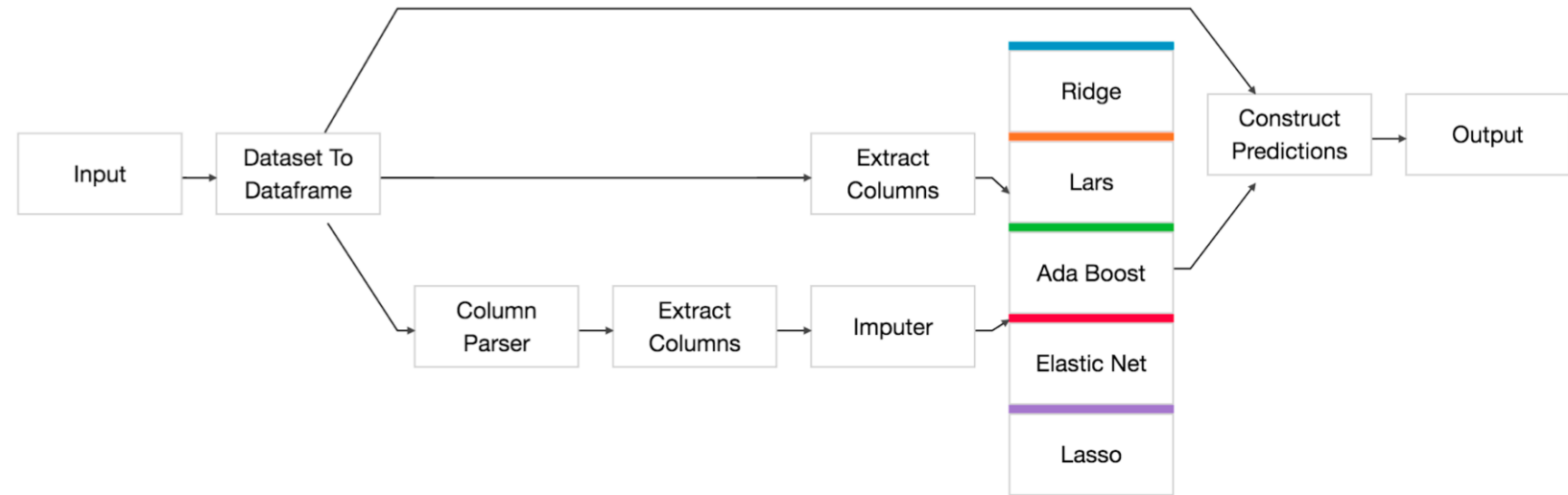


Case Study 2: Exploring AutoML Approaches

- The AlphaD3M¹ developer set out to compare how six D3M systems – denoted by A, B, C, D, E, F – performed for a regression task using the cps 85 wages dataset.
 - The systems output a total of generated 114 pipelines after 1 hour.
 - System A obtained the best performance followed by System B, System C, System D, and System E with 20.28, 20.29, 20.68, 21.46 and 21.48 mean squared error, respectively.
- Using the Pipeline Comparison View, the developer could also easily **see noticeable differences** in the **strategies** used by the AutoML systems to construct the pipelines.

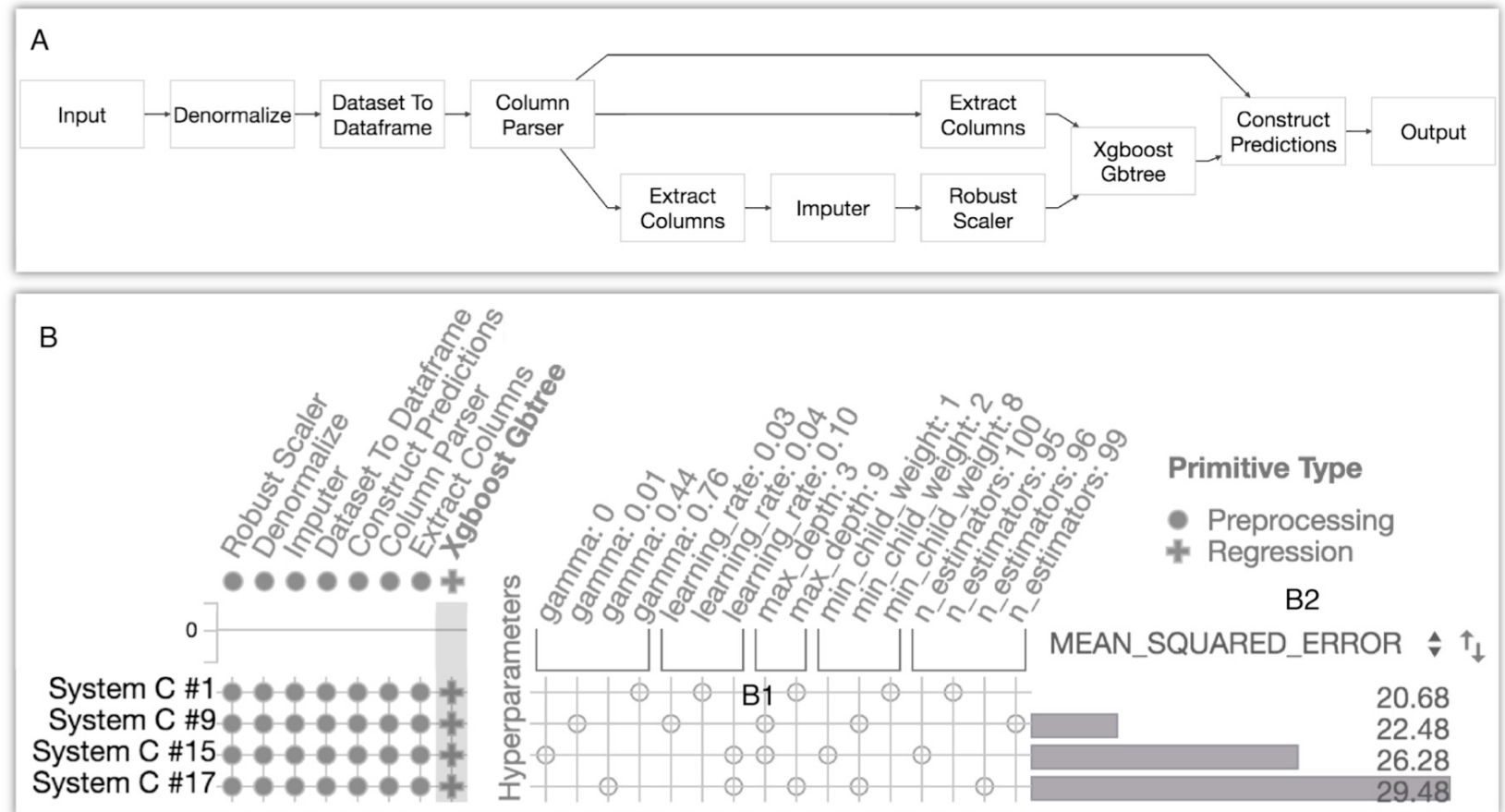
¹ AlphaD3M is an AutoML system based on reinforcement learning that uses a grammar (set of primitives' patterns) to reduce the search space of pipelines.

A visual comparison of pipelines produced by System D (6)



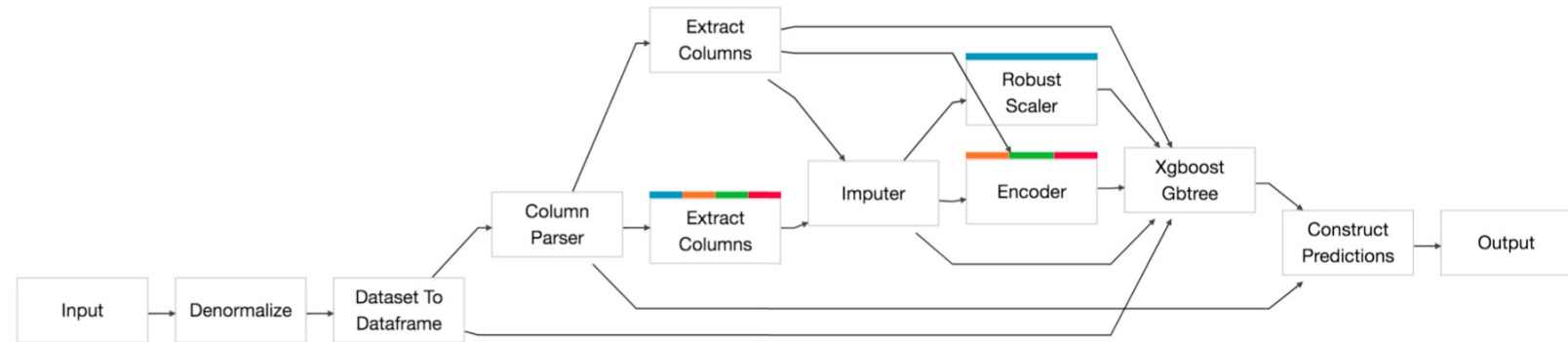
- **Suggests** that it fixes the pipeline structure and **tries multiple regression algorithms**

A visual comparison of pipelines produced by System C (7)



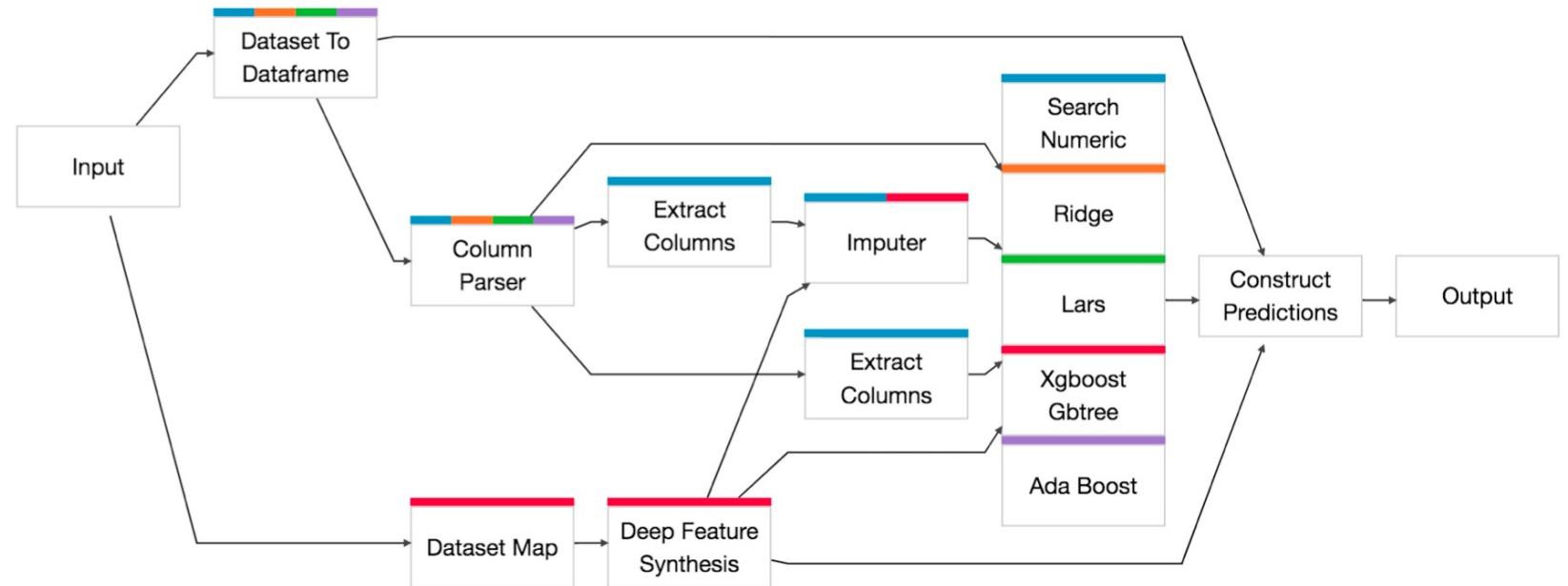
- System C produced four pipelines with the **same graph structure**. Even though these pipelines have identical structures, the **hyperparameters values** for the Xgboost Gbtree primitive **are different** (B1), and this **results in different scores** for the pipelines (B2).
- This pattern suggests that System C **tunes the hyperparameters values** after it derives the pipeline structure.

A visual comparison of pipelines produced by System C (8)



- A comparison of pipelines produced by System C indicates that, for a fixed regression algorithm (Xgboost), it **searches for alternative sequences** of preprocessing primitives.

A visual comparison of pipelines produced by System A (9)



- A comparison of pipelines produced by System A shows that these pipelines **vary both in structure and in the primitives** used, **suggesting** that it performs a **broad search** which considers **multiple preprocessing sequences** and **different regression algorithms**.

Demo

Thank you!

A thick, hand-drawn orange line that spans the width of the text above it, positioned below the 'Thank you!' message.