(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 4 439 399 A8**

(12) **CORRECTED EUROPEAN PATENT APPLICATION**

(15) Correction information:
Corrected version no 1 (W1 A2)
Corrections, see
Bibliography
Remarks

(48) Corrigendum issued on:
13.11.2024 Bulletin 2024/46

(43) Date of publication:
02.10.2024 Bulletin 2024/40

(21) Application number: **24158298.0**

(22) Date of filing: **19.02.2024**

(51) International Patent Classification (IPC):
G06N 3/12 *(2023.01)*

(52) Cooperative Patent Classification (CPC):
G06N 3/12

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC ME MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA
Designated Validation States:
GE KH MA MD TN

(30) Priority: **31.03.2023 US 202318194486**

(71) Applicant: **FUJITSU LIMITED**
Kawasaki-shi, Kanagawa 211-8588 (JP)

(72) Inventors:
- **Chen, Wei-Peng**
Fremont, CA (US)
- **Hasegawa, Sou**
Mountain View, CA (US)
- **Liu, Lei**
San Jose, CA (US)
- **Bahrami, Mehdi**
San Jose, CA (US)

(74) Representative: **Hoffmann Eitle**
Patent- und Rechtsanwälte PartmbB
Arabellastraße 30
81925 München (DE)

Remarks:
Claim 20 is deemed to be abandoned due to non-payment of the claims fee (Rule 45(3) EPC).

(54) SYNTHESIZING ML PIPELINES FOR AUTOMATED PIPELINE RECOMMENDATIONS

(57) According to an aspect of an embodiment, operations include receiving data comprising tabular datasets and code files. The operations further include generating a task specification corresponding to each dataset and determining data type information for features of each dataset. The operations further include extracting a plurality of API methods from the code files and generating an ML pipeline based on the data type information and the task specification. The operations further include obtaining variations of the ML pipeline based on options associated with at least one ML component and generating a database of pipelines based on the ML pipeline and the variations. The operations further include selecting candidate ML pipelines from the database based on an optimization approach and executing the candidate ML pipelines to evaluate a performance of each candidate pipeline on test data. The operations further include obtaining a training corpus of ML pipelines for pipeline recommendation.

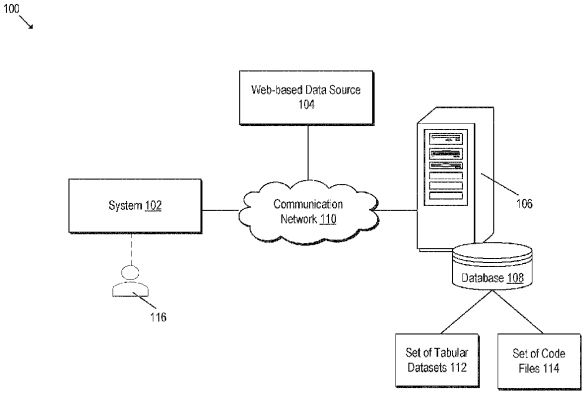

FIG. 1

EP 4 439 399 A8