



(51) International Patent Classification:

G06N 3/08 (2023.01) G06N 20/00 (2019.01)
G06Q 20/40 (2012.01) G06F 16/28 (2019.01)
G06F 16/901 (2019.01)

(21) International Application Number:

PCT/US2023/023509

(22) International Filing Date:

25 May 2023 (25.05.2023)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/345,599 25 May 2022 (25.05.2022) US

(71) Applicant: **VISA INTERNATIONAL SERVICE ASSOCIATION** [US/US]; P.O. Box 8999, San Francisco, California 94128 (US).

(72) Inventors: **ZHOU, Kaixiong**; 6100 Main Street, Houston, Texas 77005 (US). **CHEN, Huiyuan**; 919 S Winchester Blvd., San Jose, California 95128 (US). **WANG, Lan**; 429 Topaz Terrace, Sunnyvale, California 94089 (US). **BENDRE, Mangesh**; 870 E El Camino Real, Apt. 404, Sunnyvale, California 94089 (US). **WANG, Fei**; 34155 Siward Dr., Fremont, California 94555 (US). **YANG, Hao**; 2046 Wendover Lane, San Jose, California 95121 (US).

(74) Agent: **PREPELKA, Nathan, J.** et al.; The Webb Law Firm, One Gateway Center, 420 Ft. Duquesne Blvd., Suite 1200, Pittsburgh, Pennsylvania 15222 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM,

(54) Title: SYSTEM, METHOD, AND COMPUTER PROGRAM PRODUCT FOR ENCODING FEATURE INTERACTIONS BASED ON TABULAR DATA USING MACHINE LEARNING

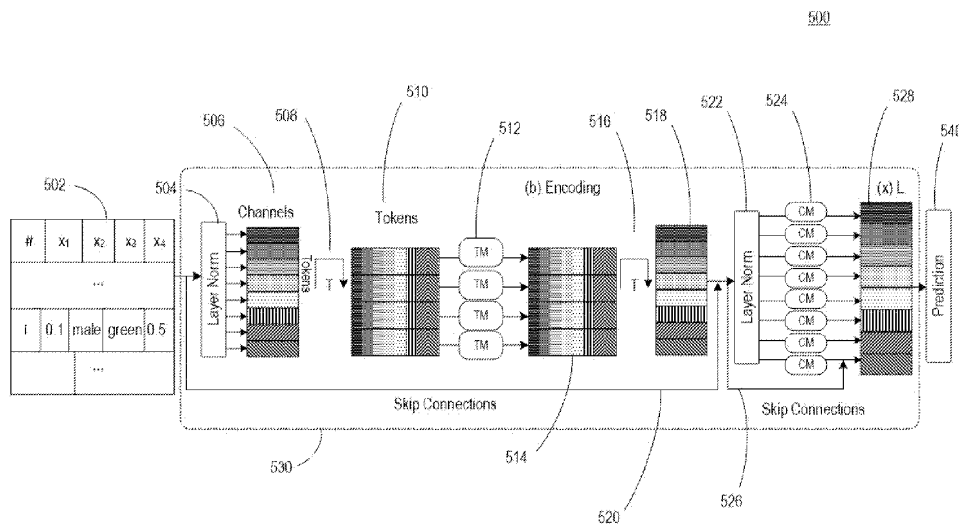


FIG. 5

(57) Abstract: Methods, systems, and computer program products are provided for encoding feature interactions based on tabular data. An exemplary method includes receiving a dataset in a tabular format including a plurality of rows and a plurality of columns. Each column is indexed to generate a position embedding matrix. Each column is grouped based on at least one tree model to generate a domain embedding matrix. An input vector is generated based on the dataset, the position embedding matrix, and the domain embedding matrix. The input vector is inputted into a first multilayer perceptron (MLP) model to generate a first output vector, which is transposed to generate a transposed vector. The transposed vector is inputted into a second MLP model to generate a second output vector, which is inputted into at least one classifier model to generate at least one prediction.

AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*
- *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

- *with international search report (Art. 21(3))*