



- (51) **International Patent Classification:**
G06N 3/04 (2006.01) *G06N 3/08* (2006.01)
- (21) **International Application Number:**
PCT/EP2021/063767
- (22) **International Filing Date:**
24 May 2021 (24.05.2021)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
63/049,873 09 July 2020 (09.07.2020) US
- (71) **Applicant: FEATURES SPACE LIMITED** [GB/GB];
Broers Building 2nd Floor, 21 JJ Thomson Avenue, Cambridge Cambridgeshire CB3 0FA (GB).
- (72) **Inventors: WONG, Kenny;** c/o Featurespace Limited, Broers Building 2nd Floor, 21 JJ Thomson Avenue, Cambridge Cambridgeshire CB3 0FA (GB). **SUTTON, David;** c/o Featurespace Limited, Broers Building 2nd Floor, 21 JJ Thomson Avenue, Cambridge Cambridgeshire CB3 0FA (GB). **BARNES-GRAHAM, Alec;** c/o Featurespace Limited, Broers Building 2nd Floor, 21 JJ Thomson Avenue, Cambridge Cambridgeshire CB3 0FA (GB). **PEREZ, Iker;** c/o Featurespace, Broers Building 2nd Floor, 21 JJ Thomson Avenue, Cambridge Cambridgeshire CB3 0FA (GB).
- (74) **Agent: KAYE, Samantha;** Taylor Vinters LLP, Merlin Place, Milton Road, Cambridge Cambridgeshire CB4 0DP (GB).
- (81) **Designated States** (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

(54) **Title:** TRAINING A MACHINE LEARNING SYSTEM FOR TRANSACTION DATA PROCESSING

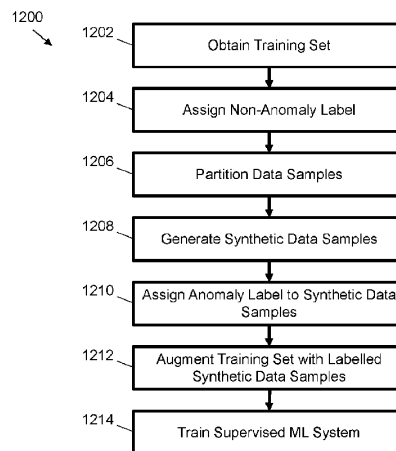


FIG. 12

(57) **Abstract:** A method of training a supervised machine learning system to detect anomalies within transaction data is described. The method includes obtaining a training set of data samples; assigning a label indicating an absence of an anomaly to unlabelled data samples in the training set; partitioning the data of the data samples in the training set into two feature sets, a first feature set representing observable features and a second feature set representing context features; generating synthetic data samples by combining features from the two feature sets that respectively relate to two different uniquely identifiable entities; assigning a label indicating a presence of an anomaly to the synthetic data samples; augmenting the training set with the synthetic data samples; and training a supervised machine learning system with the augmented training set and the assigned labels.

[Continued on next page]

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, 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, GH, GM, KE, LR, LS, MW, MZ, NA, RW, 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, 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).

Published:

— with international search report (Art. 21(3))