08/25/2017

Quality\_Control\_Data.xlsx contains 28409 rows, within which, 25188 unique rows.

Randomly chose 101 rows to QC:

'17300', '174255', '138539', '274621', '275590', '16428', '213722', '31988', '307272', '263849', '14900', '375940', '244415', '200412', '52396', '1276146', '12455', '52371', '3037376', '237527', '299661', '139216', '146559', '255086', '38768', '118719', '102381', '94450', '171849', '1349442', '1285033', '13415', '73074', '94455', '440815', '426863', '150263', '5892935', '6434069', '174159', '844750', '14996', '288436', '492465', '231050', '1414902', '130101', '228669', '311910', '18812', '135727', '127041', '288709', '318831', '242526', '12913', '311898', '36750', '190413', '489875', '228820', '279557', '271163', '150010', '174586', '3042694', '29571', '177661', '5253925', '17338', '6730891', '14445', '263539', '203359', '29836', '140076', '1060690', '11084', '72237', '174898',

'18873', '193044', '132584', '73116', '257411', '123094', '323169', '26836', '4946895', '28887', '146538', '747049', '3709488', '170136', '228818', '6574800', '8833', '199258', '304140', '271175', ‘28382’

Impression:

* High accuracy, but high false negative.

False negative examples:

* No mapping of company names: AMERICAN HEYER-SCHULTE 🡪 mentor; PIP 🡪 POLY IMPLANT PROTHESE
* tense: ruptured 🡪 rupture; deflation 🡪 deflated; leaking 🡪 leak; flattening 🡪 deflated
* “pain” was not captured as “breast pain” because “breast” is not present in the text. However, “chest pain” (5892935) should not be considered as “breast pain”?
* Hyphen: Gel-Filled 🡪 Gel
* Plural: GELS 🡪 Gel (288436); CAPSULAR CONTRCTURES 🡪 capsular contracture
* BAKER IV CAPSULE 🡪 capsular contracture
* General sickness

False positive examples:

* Negation. Positive have no CD30, and ALK context. (14996) NO LEFT BREAST MASS WAS FOUND (26836)
* Saline and silicone co-exist. (492465)
* WITHOUT EVIDENCE OF LEAKAGE OR RUPTURE (150010)
* Unrelated text in the event description (6730891). The second record is cleaner.

Need to add: siltext