

Predicting mental disorders among employees using AI- based methods

Pekka Varje, Ilkka Kivimäki, Olli Haavisto, Jussi Turtiainen,
Tiina Kalliomäki-Levanto, Aki Koskinen & Ari Väänänen

WORK2021: Work, wellbeing and health 19.8.2021

Predicting mental disorders among employees using AI-based methods

- A 2.5-year project aiming at developing novel data-driven approaches for mental health research and creating applications that support occupational health care and early prevention
- Conducted by: Finnish Institute of Occupational Health
- Funded by: The Finnish Work Environment Fund
- Personnel
 - Project leader: Research Manager Pekka Varje
 - Assisting project leader: Research Professor Ari Väänänen
 - Researchers: Specialist Researcher Tiina Kalliomäki-Levanto, Researcher Jussi Turttiainen, Senior Specialist Aki Koskinen, Data Scientist Ilkka Kivimäki, Data Scientist Olli Haavisto



Työsuojelurahasto
Arbetarskyddsfonden
The Finnish Work Environment Fund

Aims

- To produce new knowledge on the background factors and development of mental disorders among employees
- To create AI-based predictive models that facilitate the early recognition mental health problems
- To improve the ability of occupational healthcare providers and employers to support early intervention

Research questions

- What are the subjective, social and work-related factors that predict mental disorders in the Finnish working population
- What kind healthcare paths do the different background factors predict?

Research material

- Material (produced by Terveystalo, the largest healthcare service company in Finland):
 - *Sante* - occupational health questionnaires (2016–2019)
 - Electronic patient records related to mental health diagnoses (2012–2019)
 - Affective, stress-related and sleep disorders and burnout (F30–F48, F51, Z73.0)
- Two samples:
 - All persons who answered to Sante-questionnaire
 - Health questionnaires are used to predict the first mental health -related diagnosis after filling the questionnaire
 - All persons with psychiatric diagnoses
 - Patient records are used to predict treatment paths and recurring diagnoses

Methods

- Structured material
 - Predictive models based on machine learning
- Textual material
 - Natural language processing (NLP)
 - Topic modeling -> coding the material for the predictive models

Predicting psychiatric diagnoses from the health questionnaire: initial results

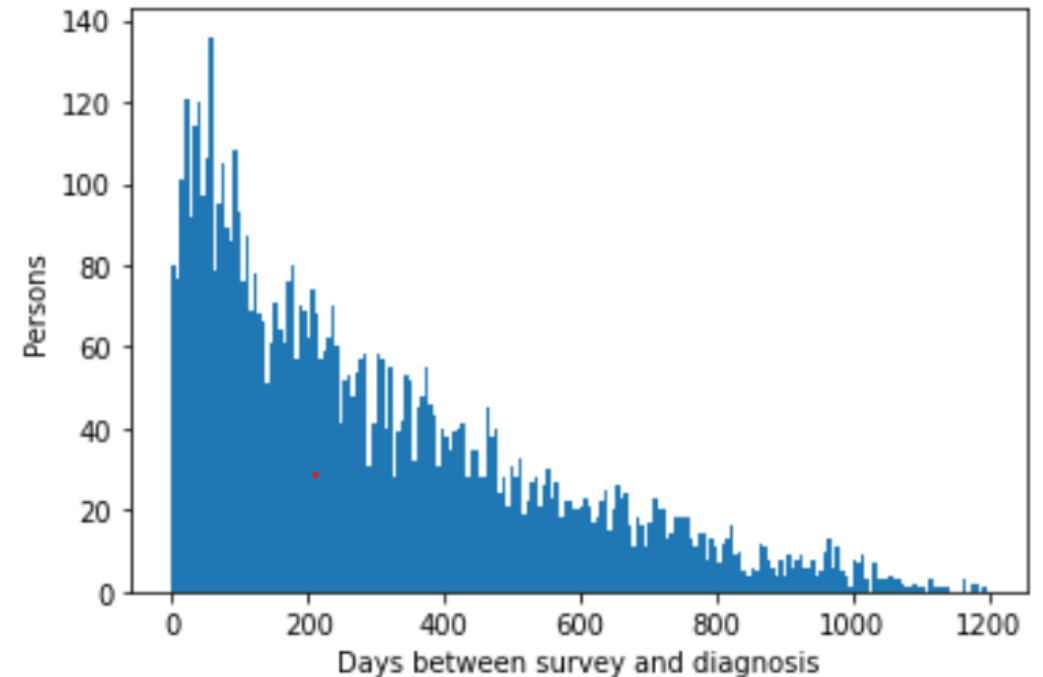


Preprocessing

- Remove persons diagnosed before survey
- Remove surveys after first diagnosis per person
- Keep only first diagnosis for each person (earliest date, randomly selected if more than one diagnosis on the same date)
- Keep only the survey closest to the diagnosis
 - Total persons left: 86842
 - With a diagnosis: 6391
 - Without a diagnosis: 80451

Preprocessing

- Filter out surveys with too long time to the diagnosis (limit = 182 days)
- Filter out too new surveys (too short follow-up period), limit = 182 days until the end of the study
 - Total persons left: 69093
 - With a diagnosis: 2622
 - Without a diagnosis: 66471



Correlations with the responses

	F3x: Mielialahäiriöt	F4x: Neuroottiset, stressiin liittyvät ja somatoformiset häiriöt	F5x: Ei-elimelliset unihäiriöt	Z73.0 Työuupumus	Ei diagnoosia
QDiabetesFamily	-0.001020	0.010814	0.004427	0.008440	-0.011642
QDiabetesYou	0.001018	-0.002165	-0.010434	-0.005745	0.008494
QDiseaseAllergy	-0.006514	-0.007913	-0.007729	-0.005598	0.013877
QDiseaseAnemia	-0.008399	-0.020823	-0.008906	-0.003156	0.023596
QDiseaseAsthma	-0.011916	-0.009489	-0.003503	0.000276	0.013262
QDiseaseCancer	0.003158	-0.001245	0.001504	0.003116	-0.002039
QDiseaseCardiovascular	0.000767	0.000521	0.003816	0.000884	-0.003080
QDiseaseDiabetes	-0.002476	0.003135	0.000120	0.003977	-0.002204
QDiseaseEpilepsy	-0.008727	0.001781	-0.003330	-0.000401	0.004295
QDiseaseHypertension	0.006184	0.009662	-0.000759	0.002722	-0.009357
QDiseaseMigraine	-0.015738	-0.027973	-0.017531	-0.013301	0.038908
QDiseaseMusculoskeletal	-0.007825	-0.009083	-0.023184	-0.005117	0.023992
QDiseaseThyroid	-0.006074	-0.022231	-0.013424	-0.010992	0.028148
QGender	0.020594	0.049494	0.031693	0.026611	-0.067070
QHearingDiagnosed	-0.004240	0.002087	-0.005893	0.000285	0.003591
QHearingWork	-0.014883	-0.005925	-0.014171	-0.007855	0.020132

QHeavyExerciseDays	-0.004927	-0.003417	-0.003938	-0.006991	0.008305
QJobContentsCreativity	-0.005889	-0.009980	-0.004844	0.006296	0.010480
QJobContentsDevelopAbilities	-0.019770	-0.019450	-0.018447	-0.023144	0.037596
QJobContentsHighSkill	-0.009876	-0.018092	0.000935	0.005322	0.014571
QJobContentsLearnNew	0.002343	0.006030	0.008317	0.007821	-0.011790
QJobDemandsExcessiveDemands	0.026576	0.030192	0.037040	0.050992	-0.065232
QJobDemandsHectic	0.009865	0.012365	0.024641	0.033982	-0.034962
QJobDemandsNotEnoughTime	0.027720	0.018379	0.030657	0.039595	-0.051104
QJobDemandsWorkingFast	0.012415	0.019244	0.020838	0.028307	-0.037141
QJobDemandsWorkingHard	0.011713	0.013612	0.025514	0.026100	-0.035139
QMedicationAntiAllergic	-0.005389	-0.006315	-0.010446	-0.007066	0.014250
QMedicationCardioVascular	0.004335	0.013473	0.000423	-0.000770	-0.011063
QMedicationDermatological	-0.004440	-0.006705	-0.006740	-0.000602	0.010424
QMedicationGastrointestinal	-0.010610	-0.014607	-0.009723	-0.012828	0.023042
QMedicationHormonal	-0.015859	-0.020027	-0.020133	-0.014837	0.035364
QMedicationMetabolic	-0.007337	-0.010254	-0.006631	-0.003535	0.014682
QMedicationMusculoskeletal	-0.012490	-0.016496	-0.020691	-0.014496	0.031821
QMedicationNeurologic	-0.013311	-0.018810	-0.015827	-0.006303	0.028938
QMedicationPulmonary	-0.016902	-0.010499	-0.007682	-0.004896	0.019629
QMediumHeavyExerciseDays	-0.004884	0.000048	0.003022	-0.014425	0.003713

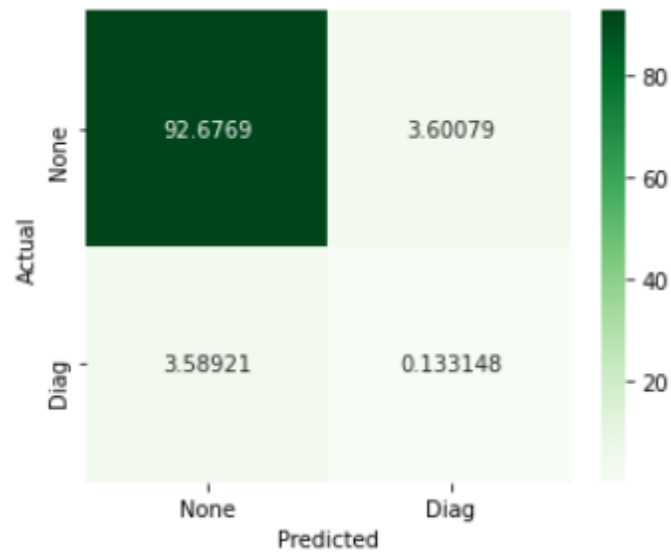
QMoodDisappearedJoy	0.119013	0.069226	0.053205	0.042871	-0.136518
QMoodEffort	0.102100	0.073948	0.060846	0.052335	-0.139691
QMoodEnjoyLife	0.077481	0.049409	0.044796	0.032470	-0.098828
QMoodHopelessFuture	0.109114	0.072018	0.053328	0.046311	-0.135377
QMoodInsomnia	0.050850	0.037130	0.111421	0.037016	-0.119171
QMoodLethargic	0.092138	0.057783	0.051139	0.037407	-0.115332
QMoodLoneliness	0.092008	0.057613	0.034076	0.022790	-0.101764
QMoodMelancholy	0.112865	0.068489	0.049015	0.040561	-0.130566
QMoodSad	0.102134	0.095082	0.056208	0.037325	-0.147903
QMoodWorthless	0.107976	0.070973	0.041966	0.038876	-0.125841
QMusculoskeletalSymptomAnkle	-0.006218	-0.005151	-0.008780	-0.000383	0.011190
QMusculoskeletalSymptomArm	-0.012296	-0.012920	-0.018542	-0.003461	0.025349
QMusculoskeletalSymptomBack	-0.022396	-0.020099	-0.024498	-0.013188	0.040149
QMusculoskeletalSymptomElbow	-0.009292	-0.005831	-0.017841	-0.006470	0.019598
QMusculoskeletalSymptomHip	-0.009092	-0.014275	-0.015276	-0.012148	0.025236
QMusculoskeletalSymptomKnee	-0.007745	-0.007957	-0.008422	0.003735	0.012520
QMusculoskeletalSymptomNeck	-0.019458	-0.025524	-0.027453	-0.019300	0.045897
QMusculoskeletalSymptomShoulder	-0.001659	-0.005702	-0.012586	-0.001284	0.012150
QMusculoskeletalSymptomWrist	-0.014123	-0.014241	-0.019482	-0.006025	0.028158

QPainOften	0.025016	0.022587	0.031156	0.006493	-0.045110
QPhysicalLimitationOutsideWork	0.025688	0.025059	0.034491	0.013426	-0.050695
QPhysicalLimitationWork	0.034163	0.039731	0.039861	0.023667	-0.069786
QRecoveryEveryDay	-0.055841	-0.050060	-0.048274	-0.027860	0.091465
QRecoveryWeekend	-0.077398	-0.063264	-0.060303	-0.044598	0.120220
QRewardingAppreciation	-0.038898	-0.043904	-0.042481	-0.043204	0.080841
QRewardingCareerDevelopment	-0.019330	-0.027203	-0.036185	-0.018644	0.051869
QRewardingIncome	-0.033559	-0.039316	-0.030463	-0.029112	0.065177
QRewardingMeaningfulness	-0.044985	-0.041743	-0.039206	-0.047656	0.081001
QSitExerciseHoursPerDay	0.010259	-0.000699	-0.002211	0.001377	-0.002701
QSleepDifficultiesFallingAsleep	0.044769	0.044586	0.085790	0.024590	-0.104060
QSleepDozeReading	0.029564	0.015217	0.013577	0.023044	-0.035788
QSleepDozeWatchingTV	0.028361	0.018224	0.021099	0.015976	-0.039978
QSleepExhaustedDaytime	0.074975	0.062789	0.083064	0.045592	-0.132272
QSleepHoursPerDay	-0.006869	0.002235	-0.034671	-0.003486	0.022039
QSleepInAStoppedCar	0.023991	0.012805	0.015769	0.003462	-0.028368
QSleepLyingDown	0.037921	0.016833	0.015535	0.016694	-0.039830
QSleepNeedPerDay	0.036809	0.037223	0.029784	0.023000	-0.063157
QSleepPassengerInACar	0.038060	0.023800	0.016303	0.020254	-0.045991
QSleepSittingAfterLunch	0.029422	0.011910	0.016421	0.012693	-0.032564
QSleepSittingAndTalking	0.030097	0.012517	0.012663	0.001207	-0.028281
QSleepSittingInactive	0.031569	0.007827	0.018829	0.001242	-0.029205
QSleepTiredDaytime	0.069745	0.058314	0.072085	0.041923	-0.119874
QSleepWakenUp	0.013395	0.019007	0.053783	0.020124	-0.054345

QSymptomAnxiety	-0.086838	-0.103485	-0.047696	-0.034366	0.141916
QSymptomConsciousness	-0.004180	-0.013394	-0.009365	-0.005638	0.017663
QSymptomHeadaches	-0.037650	-0.040061	-0.032777	-0.019017	0.066197
QSymptomJoints	-0.009211	-0.010916	-0.014495	-0.006632	0.021175
QSymptomLowerLimbs	-0.014444	-0.014250	-0.018084	-0.009257	0.028279
QSymptomUpperLimbs	-0.015580	-0.016459	-0.015798	-0.013053	0.029866
QSymptomVertigo	-0.026647	-0.038229	-0.025998	-0.017921	0.056348
QVisionProblems	-0.011858	-0.017679	-0.017673	-0.010107	0.029570
QVisionTest	-0.005697	0.001411	-0.009345	0.000126	0.006667
QWalkExerciseDays	-0.005701	0.008709	0.001354	-0.000174	-0.004439
QWellbeingEnjoyLife	-0.080982	-0.049689	-0.043223	-0.030958	0.099151
QWellbeingEnjoyWork	-0.052550	-0.043650	-0.041494	-0.046725	0.086440
QWellbeingFamilyProblems	-0.042886	-0.061479	-0.035765	-0.010605	0.082697
QWellbeingFamilyTime	-0.045958	-0.038746	-0.033371	-0.032990	0.072385
QWellbeingSqueezedWork	-0.052874	-0.049008	-0.051573	-0.059552	0.099195
QWellbeingStressed	-0.085895	-0.079888	-0.086620	-0.056851	0.153218
QWellbeingWorkBullying	-0.019390	-0.038759	-0.026315	-0.023418	0.055325
QWellbeingWorkCommunity	-0.030613	-0.037501	-0.023962	-0.030830	0.059421
QWellbeingWorkSchedule	-0.023748	-0.027947	-0.022694	-0.032247	0.049709
QWellbeingWorkSuperior	-0.024041	-0.019374	-0.028360	-0.032582	0.047275
QWellbeingWorkTime	-0.022498	-0.027532	-0.020420	-0.021476	0.044980
QWellbeingWorkVictim	-0.021218	-0.046556	-0.030085	-0.027520	0.064596
QWorkAbilityPrognosis	0.052261	0.042260	0.044653	0.047430	-0.087359

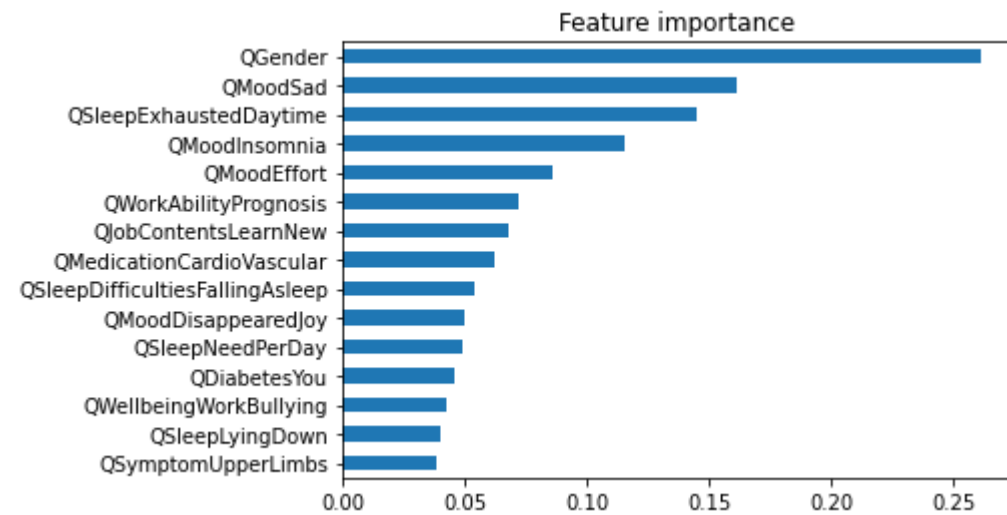
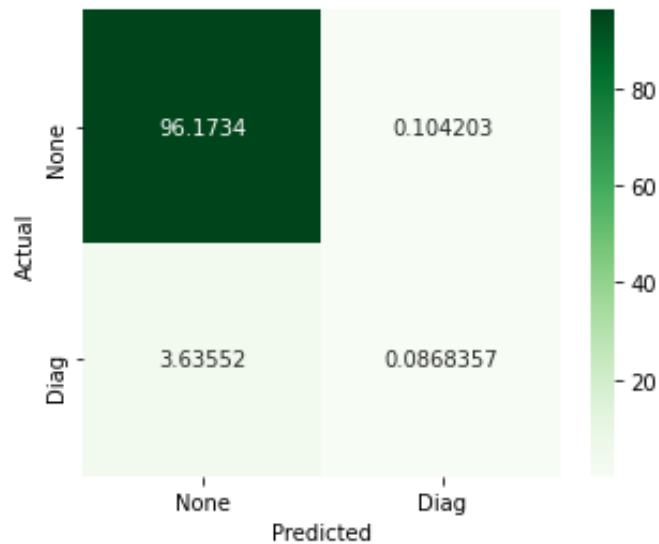
Classifying with random classifier (baseline)

	precision	recall	f1-score	support
0	0.96	0.96	0.96	16631
1	0.04	0.04	0.04	643
accuracy			0.93	17274
macro avg	0.50	0.50	0.50	17274
weighted avg	0.93	0.93	0.93	17274



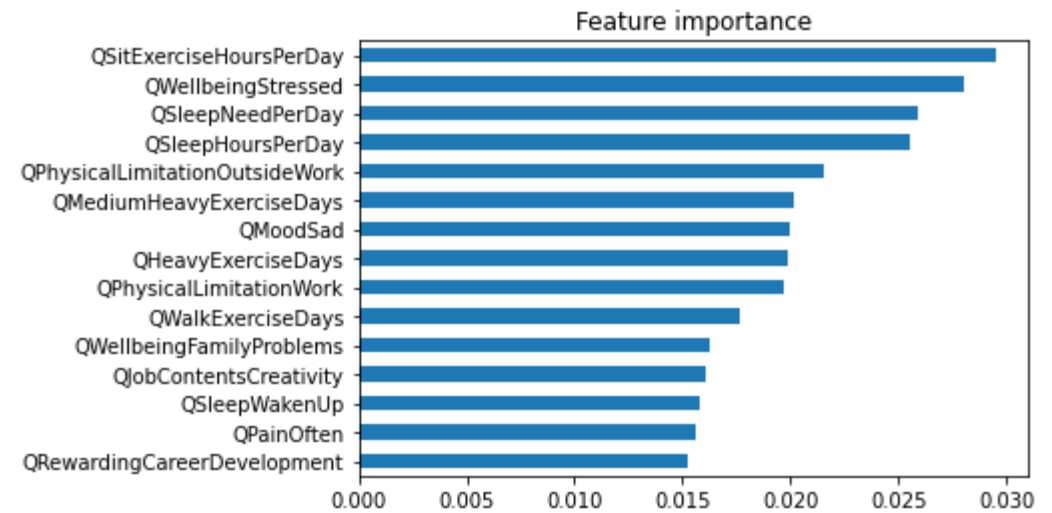
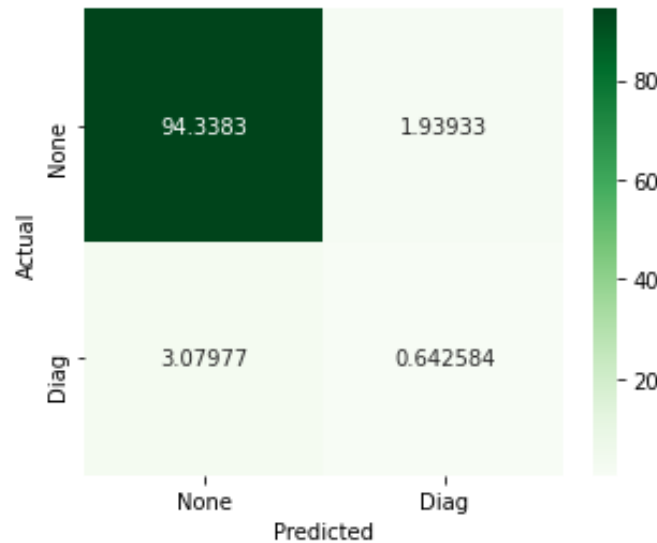
Classifying with logistic regression – responses

	precision	recall	f1-score	support
0	0.96	1.00	0.98	16631
1	0.45	0.02	0.04	643
accuracy			0.96	17274
macro avg	0.71	0.51	0.51	17274
weighted avg	0.94	0.96	0.95	17274



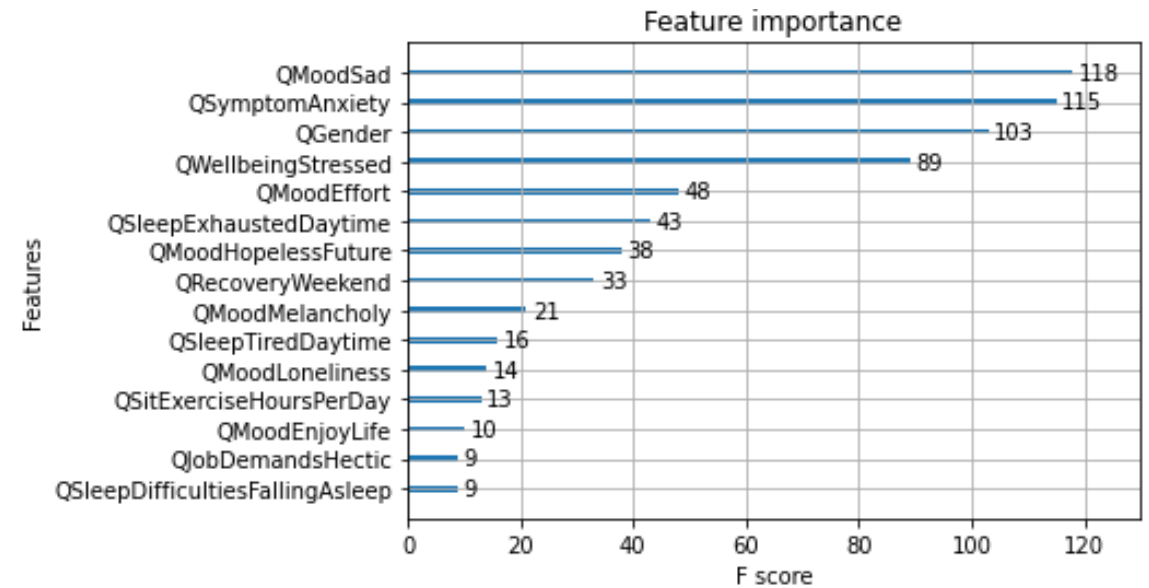
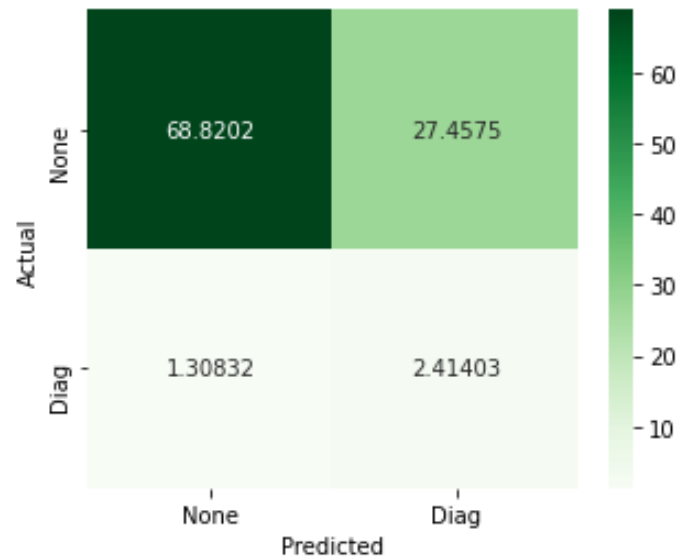
Classifying with Random forest – responses

	precision	recall	f1-score	support
0	0.97	0.98	0.97	16631
1	0.25	0.17	0.20	643
accuracy			0.95	17274
macro avg	0.61	0.58	0.59	17274
weighted avg	0.94	0.95	0.95	17274



Classifying with XGBoost – responses

	precision	recall	f1-score	support
0	0.98	0.71	0.83	16631
1	0.08	0.65	0.14	643
accuracy			0.71	17274
macro avg	0.53	0.68	0.49	17274
weighted avg	0.95	0.71	0.80	17274



Finnish Institute of
Occupational Health



Työsuojelurahasto
Arbetarskyddsfonden
The Finnish Work Environment Fund

Thank you!

pekka.varje@ttl.fi

