Feature Forward: Standardizing and Monitoring Epic Data for Smarter Models

Scott Nelson, PharmD, MS
Department of Biomedical Informatics, Vanderbilt University

PURPOSE

- Develop a VUMC-wide “feature library” for predictive modeling using Epic data sources
- Implement a framework for monitoring feature-level drift

METHODS

Meta data captured

REDCap®

ETL into databricks

Visualization in Tableau/VAMOS

tableau

VAMOS Collaborative
Vigilant Governance. Trusted AI.

https://tableau.app.vumc.org/#/site/VCLIC/workbooks/18532/views

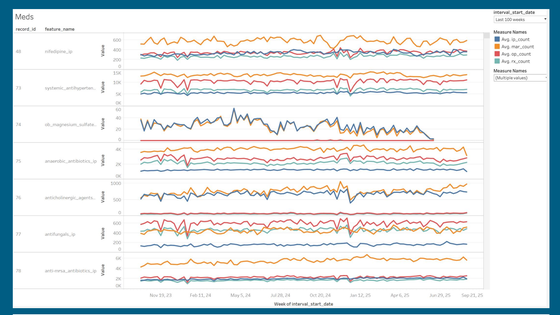
VCLIC
VANDERBILT CLINICAL INFORMATICS CENTER

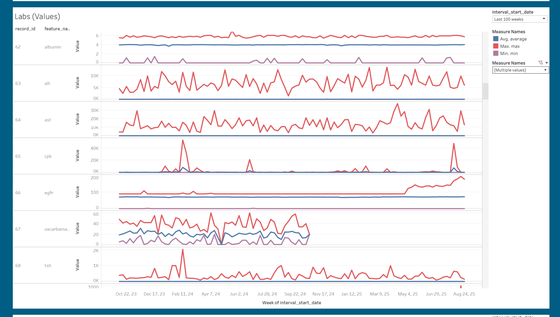
A standardized Epic feature catalog with embedded drift monitoring offers a promising approach to improving model robustness and reproducibility

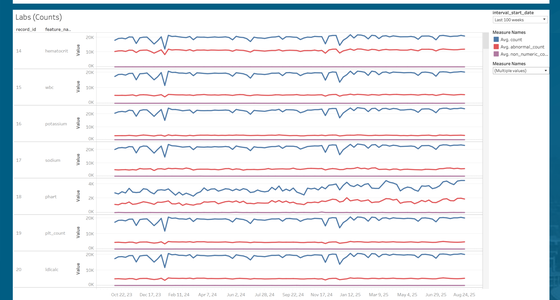
Summary | Meds | Meds (Missing) | Labs (Values) | Labs (Counts) | Labs (Missing) | Problems | Problems (Missing) | Flowsheet (Values) | Flowsheet (Counts)

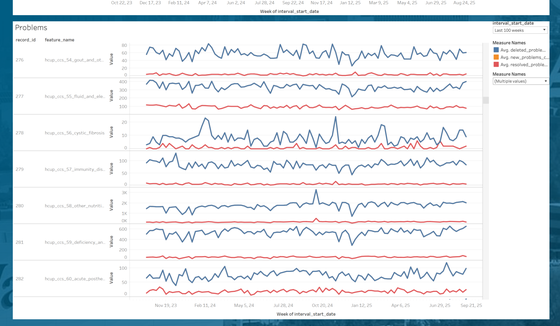
Summary

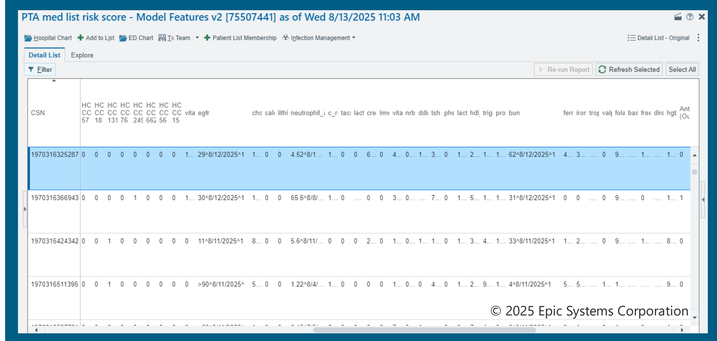
feature_cat..	feature_so..	feature_name	feature_description	epic_cer_id	epic_paf_id	feature_val..	feature_source_ids	Counts
Diagnosis	EDG	hcup_ccs_52_nutritional..	HCUP CCS 52-NUTRITIONAL DEFICIENC..	797807	108231	1750052		44,220
		hcup_ccs_79_parkinsons..	HCUP CCS 79-PARKINSONS DISEASE	797823	108247	1750079		2,732
		hcup_ccs_98_essential_h..	HCUP CCS 98-ESSENTIAL HYPERTENSI..	797842	108266	1750098		105,933
		hcup_ccs_100_acute_my..	HCUP CCS 100-ACUTE MYOCARDIAL IN..	797844	108268	1750100		6,808
		hcup_ccs_128_asthma	HCUP CCS 128-ASTHMA	797872	108296	1750128		27,809
		hcup_ccs_202_rheumatoi..	HCUP CCS 202-RHEUMATOID ARTHRITI..	797946	108370	1750202		5,700
		hcup_ccs_203_osteoarth..	HCUP CCS 203-OSTEOARTHRITIS	797947	108371	1750203		28,180
Lab	LRR	albumin	Albumin level (ALBUMIN,POCALB) lab	799022	108577	101811	ALBUMIN,POCALB	1,210,325
		alt	ALT level (ALT,POCALT,SGPTALT) lab	799023	108578	142561	ALT,POCALT,SGPTALT	1,226,726
		ast	AST level (AST,POCAST,SGOTAST) lab	799024	108579	142563	AST,POCAST,SGOTAST	1,225,687
		bun	BUN level lab	799006	108560	2103333267	BUN,BUN2,BUN3	2,050,573
		calcium	Calcium level (CALCIUM) lab	799010	108564	2100000780	CALCIUM,CALCIUM2	2,035,294
		free_t4	Thyroxine free (T4FREE) lab	799095	108649	2100000760	T4FREE	146,537
		hgb	Hemoglobin (HGB) lab	799038	108592	2103330024	HGB	0
		hgb_poct	Hemoglobin POCT (POCHGB) lab	800820	108666	5853330032	POCHGB	61
		hgba1c	Hemoglobin a1c level (HGBA1C) lab	799073	108627	2103330032	HGBA1C	2,158,042
		inr	International normalization ratio (INR)..	799061	108615		INR	410,101
		lithium_level	Lithium lab result (LITHIUMLEVEL) lab	799013	108568		LITHIUMLEVEL	3,410
		lmwh_anti_xa	LMWH anti-xa level (LMWHEPARIN) lab	799981	108646		LMWHEPARIN	15,861
Medications	ORD	acyclovir_iv_op	Acyclovir IV (Outpatient)	798322	107734	4080200086		3
		all_systemic_antimicrobi..	All systemic antimicrobials (Outpatient)	798296	107707	4080200074		1,555,849
		aminoglycosides_ip	Aminoglycosides (Inpatient)	798625	107510	100888		12,547
		amoxicillin_op	Amoxicillin (Outpatient)	798360	107783	4080200096		144,810
		antiarrhythmics_op	Antiarrhythmics (Outpatient)	798427	107746	142548		0
		anticonvulsants_ip	Anticonvulsants (Inpatient)	798601	107684	113575		466,713
		antiplatelet_minus_aspri..	Antiplatelet (minus aspirin) (Outpatie..	798417	107770	120240		63,781
		aztreonam_iv_ip	Aztreonam IV (Inpatient)	798487	107520	140819		10
		benzodiazepines_op	Benzodiazepines (Outpatient)	798424	107773	108290		481,321
		enoxaparin_ip	Enoxaparin (Inpatient)	798505	107539	4080100006		93,078
		immunosuppressive_age..	Immunosuppressive agents (Outpatie..	798437	107842	113577		172,571
		medd_90	1 if MEDD is >=90	801301	108480			0
		missing_unknown_medic..	Missing or Unknown Medication (Outp..	798798	107878	142557		0
Vitals	FLO	bp_diastolic	Diastolic blood pressure (mmHg)				5	16,164,014
		bp_systolic	Systolic blood pressure (mmHg)				5	16,164,014
		height_m	Height (meters)		4312		11	3,715,083

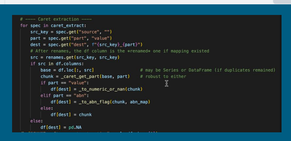




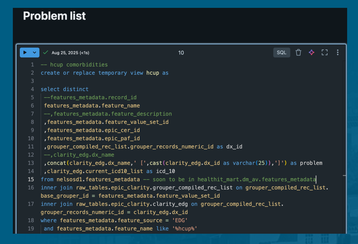








Scan for full text



Results

- *845 Features added*
  - *809 CER rules*
  - *814 PAF columns*
  - *694 VCG records*

- ***Feature groups***
  - *ADT: 11*
  - *Demographics: 42*
  - *Diagnosis: 270*
  - *Lab results: 119*
  - *Medications: 383*
  - *Flowsheets: 11*
  - *Vitals: 9*

- *Used to build 2 predictive models*
  - *SQL code*
  - *Python script / cleaning*

Presenter:
Scott Nelson

Scott.Nelson@vumc.org

@scottnelsonrx

VANDERBILT UNIVERSITY
MEDICAL CENTER

Department of
Biomedical Informatics