

HYPERLOOP FEATURE ENHANCEMENT PROPOSAL

for Machine Learning model generated features

History			
NR	Data	Autor	Ver
1	01.02.2017	A.I. DAMIAN	Draft 1
2	02.02.2017	A.I. DAMIAN	Draft 2

Contents

1	Features derived from Machine Learning models for client brand preference	1
2	Features derived from Machine Learning models for client product categories preference	2
3	Features derived from Machine Learning models for client price range (level) preference	5

1 Features derived from Machine Learning models for client brand preference

Obs.: all 17 proposed features have descriptive names instead of short cryptic codes due to the need for high self-explain ability)

Name	Description	Notes
BP01_APIVITA	Client brand preference for APIVITA	Real value score [0.00-1.00]
BP02_OCCITANE	Client brand preference for L'OCCITANE	Real value score [0.00-1.00]
BP03_VICHY	Client brand preference for VICHY	Real value score [0.00-1.00]
BP04_DR_HART	Client brand preference for DR HART	Real value score [0.00-1.00]
BP05_NUROFEN	Client brand preference for NUROFEN	Real value score [0.00-1.00]
BP06_AVENE	Client brand preference for AVENE	Real value score [0.00-1.00]
BP06_BIODERMA	Client brand preference for BIODERMA	Real value score [0.00-1.00]
BP07_LA_ROCHE	Client brand preference for LA ROCHE POSAY	Real value score [0.00-1.00]
BP08_ERBOLARIO	Client brand preference for L' ERBOLARIO	Real value score [0.00-1.00]
BP09_PARASINUS	Client brand preference for PARASINUS	Real value score [0.00-1.00]
BP10_NESTLE	Client brand preference for NESTLE	Real value score [0.00-1.00]
BP11_OXYANCE	Client brand preference for OXYANCE	Real value score [0.00-1.00]
BP12_DUCRAY	Client brand preference for DUCRAY	Real value score [0.00-1.00]
BP13_SCHOLL	Client brand preference for SCHOLL	Real value score [0.00-1.00]
BP14_ASPENTER	Client brand preference for ASPENTER	Real value score [0.00-1.00]
BP15_TRUSSA	Client brand preference for TRUSSA	Real value score [0.00-1.00]
BP16_SREPSILS	Client brand preference for SREPSILS	Real value score [0.00-1.00]
BP17_MELVITA	Client brand preference for MELVITA	Real value score [0.00-1.00]

Client brand preferences features (BP_<BRAND>) define the actual inclination (disposition) that a particular client has towards buying a certain brand. The real value score is in range 0.00 up to 1.00 where a value of 0 means that the client has never bought that particular brand, a value of 1.00 means

a potential client that buys only that particular brand and a value around 0.5 means the client buys about half of his/hers products from that particular brand.

The brand preference features can be used directly or in a model (ranging from simple statistical models to complex machine learning models) for the following purposes:

- a) Customer behavior segmentation
- b) Market-Basket based segmentation or individual market basket inference
- c) Products/Buy recommendation
- d) New product campaigns (for certain brands) or other kind of campaigns

2 Features derived from Machine Learning models for client product categories preference

(obs.: all 60 features have descriptive names instead of cryptic codes due to the need for high self-explain ability)

Name	Description	Notes
CP01_BABY	Client preference for products tagged as BABY products (category)	Real value score [0.00-1.00]
CP02_COSMETICE	Client preference for products tagged as COSMETICE products (category)	Real value score [0.00-1.00]
CP03_DERMOCOSM	Client preference for products tagged as DERMOCOSMETICE products (category)	Real value score [0.00-1.00]
CP04_PHARMA	Client preference for products tagged as PHARMA products (category)	Real value score [0.00-1.00]
CP05_SPECIALE	Client preference for products tagged as SPECIALE products (category)	Real value score [0.00-1.00]
CP06_OTC	Client preference for products tagged as OTC products (category)	Real value score [0.00-1.00]
CP07_MEDICAT_ALTERN	Client preference for products tagged as MEDICATIE_ALTERNATIVA products (category/ subcategory)	Real value score [0.00-1.00]
CP08_BARBATI	Client preference for products tagged as BARBATI products (category/ subcategory)	Real value score [0.00-1.00]
CP09_CADOURI	Client preference for products tagged as CADOURI products (category/ subcategory)	Real value score [0.00-1.00]
CP10_COPII	Client preference for products tagged as COPII products (category/ subcategory)	Real value score [0.00-1.00]
CP11_FRUM_PRENAT	Client preference for products tagged as FRUMUSETE_PRE_NATAL products (category/ subcategory)	Real value score [0.00-1.00]
CP12_FRUMUSETE	Client preference for products tagged as FRUMUSETE products (category/ subcategory)	Real value score [0.00-1.00]
CP13_HOMEOPATIE	Client preference for products tagged as HOMEOPATIE products (category/ subcategory)	Real value score [0.00-1.00]
CP14_PROMO	Client preference for products tagged as PROMO products (category/ subcategory)	Real value score [0.00-1.00]
CP15_R_APARAT_RESP	Client preference for products tagged as R_APARAT_RESPIRATOR products (category/ subcategory)	Real value score [0.00-1.00]
CP16_S_ORGANE_SENZ	Client preference for products tagged as S_ORGANE_SENZITIVE products (category/ subcategory)	Real value score [0.00-1.00]

CP17_SANATATE	Client preference for products tagged as SANATATE products (category/ subcategory)	Real value score [0.00-1.00]
CP18_STYLIST	Client preference for products tagged as STYLIST products (category/ subcategory)	Real value score [0.00-1.00]
CP19_INGRIJIRE_CORP	Client preference for products tagged as INGRIJIRE_CORP products (category/ subcategory)	Real value score [0.00-1.00]
CP20_INGRIJIRE_PERS	Client preference for products tagged as INGRIJIRE_PERSONALA products (category/ subcategory)	Real value score [0.00-1.00]
CP21_SANGE_ORG_HP	Client preference for products tagged as B_SANGE_SI_ORGANE_HEMATOPOETICE products (category/ subcategory)	Real value score [0.00-1.00]
CP22_PENTRU_EA	Client preference for products tagged as PENTRU_EA products (category/ subcategory)	Real value score [0.00-1.00]
CP23_SAMPON_USCAT	Client preference for products tagged as SAMPON_USCAT products (category/ subcategory)	Real value score [0.00-1.00]
CP24_SIS_MUSC_SCHEL	Client preference for products tagged as M_SISTEMUL_MUSCULO_SCHELETIC products (category/ subcategory)	Real value score [0.00-1.00]
CP25_INGRIJIRE_INTIMA	Client preference for products tagged as INGRIJIRE_INTIMA products (category/ subcategory)	Real value score [0.00-1.00]
CP26_INGRIJIRE_FATA	Client preference for products tagged as INGRIJIRE_FATA products (category/ subcategory)	Real value score [0.00-1.00]
CP27_CADOURI_MIXTE	Client preference for products tagged as CADOURI_MIXTE products (category/ subcategory)	Real value score [0.00-1.00]
CP28_N06_PSIHOAN	Client preference for products tagged as N06_PSIHOANALEPTICE products (category/ subcategory)	Real value score [0.00-1.00]
CP29_ANTIIST_UZ_SIST	Client preference for products tagged as R06_ANTIISTAMINICE_DE_UZ_SISTEMIC products (category/ subcategory)	Real value score [0.00-1.00]
CP30_INCALT_ACCESOR	Client preference for products tagged as INCALTAMINTE_SI_ACCESORII products (category/ subcategory)	Real value score [0.00-1.00]
CP31_PROT_SEX	Client preference for products tagged as PROTECTIE_SEXUALA products (category/ subcategory)	Real value score [0.00-1.00]
CP32_TRACT_DIG_MET	Client preference for products tagged as A_TRACTUL_DIGESTIV_SI_METABOLISM products (category/ subcategory)	Real value score [0.00-1.00]
CP33_BARBIERIT	Client preference for products tagged as BARBIERIT products (category/ subcategory)	Real value score [0.00-1.00]
CP34_CARDIOVASC	Client preference for products tagged as C_SISTEM_CARDIOVASCULAR products (category/ subcategory)	Real value score [0.00-1.00]
CP35_BAIE	Client preference for products tagged as BAIE products (category/ subcategory)	Real value score [0.00-1.00]
CP36_PARFUM	Client preference for products tagged as PARFUMERIE products (category/ subcategory)	Real value score [0.00-1.00]
CP37_INGRIJ_PICIO	Client preference for products tagged as INGRIJIRE_PICIOARE products (category/ subcategory)	Real value score [0.00-1.00]

CP38_INGRIJIRE_SOL	Client preference for products tagged as INGRIJIRE_SOLARA products (category/ subcategory)	Real value score [0.00-1.00]
CP39_DISP_MEDICAL	Client preference for products tagged as DISPOZITIVE_MEDICALE products (category/ subcategory)	Real value score [0.00-1.00]
CP40_JUCARII	Client preference for products tagged as JUCARII products (category/ subcategory)	Real value score [0.00-1.00]
CP41_PENTRU_COPII	Client preference for products tagged as PENTRU_COPII products (category/ subcategory)	Real value score [0.00-1.00]
CP42_INGRIJIRE_PAR	Client preference for products tagged as INGRIJIREA_PARULUI products (category/ subcategory)	Real value score [0.00-1.00]
CP43_OCHELARI	Client preference for products tagged as OCHELARI products (category/ subcategory)	Real value score [0.00-1.00]
CP44_SIST_NERVOS	Client preference for products tagged as N_SISTEMUL_NERVOS products (category/ subcategory)	Real value score [0.00-1.00]
CP45_MAKE_UP	Client preference for products tagged as MAKE_UP products (category/ subcategory)	Real value score [0.00-1.00]
CP46_ANALGEZICE	Client preference for products tagged as N02_ANALGEZICE products (category/ subcategory)	Real value score [0.00-1.00]
CP47_CASA	Client preference for products tagged as ACCESORII_CASA products (category/ subcategory)	Real value score [0.00-1.00]
CP48_JUCARII	Client preference for products tagged as JUCARII products (category/ subcategory)	Real value score [0.00-1.00]
CP49_ANTIINFECTII	Client preference for products tagged as J_ANTIINFECTIOASE_DE_UZ_SISTEMIC products (category/ subcategory)	Real value score [0.00-1.00]
CP50_INGR_DENTARA	Client preference for products tagged as INGRIJIRE_DENTARA products (category/ subcategory)	Real value score [0.00-1.00]
CP51_DIETETICE1	Client preference for products tagged as V06_DIETETICE products (category/ subcategory)	Real value score [0.00-1.00]
CP52_NETERAPEUTICE	Client preference for products tagged as V07_NETERAPEUTICE products (category/ subcategory)	Real value score [0.00-1.00]
CP53_OFTALMOLOGIC	Client preference for products tagged as S01_PRODUSE_OFTALMOLOGICE products (category/ subcategory)	Real value score [0.00-1.00]
CP54_ANESTEZICE	Client preference for products tagged as N01_ANESTEZICE products (category/ subcategory)	Real value score [0.00-1.00]
CP55_PREP_NAZALE	Client preference for products tagged as R01_PREP_NAZALE products (category/ subcategory)	Real value score [0.00-1.00]
CP56_TOALETA_ZILNIC	Client preference for products tagged as TOALETA_ZILNICA products (category/ subcategory)	Real value score [0.00-1.00]
CP57_HRANA	Client preference for products tagged as HRANA products (category/ subcategory)	Real value score [0.00-1.00]
CP58_PSIHOLEPTICE	Client preference for products tagged as N05_PSIHOLEPTICE products (category/ subcategory)	Real value score [0.00-1.00]
CP59_TINCT_HOMEEO	Client preference for products tagged as XRNT_TINCTURI_HOMEOPATE products (category/ subcategory)	Real value score [0.00-1.00]
CP60_ANESTEZICE	Client preference for products tagged as ANESTEZICE products (category/ subcategory)	Real value score [0.00-1.00]

Client category preferences features (CP_<CATEG>) define the actual inclination (disposition) that a particular client has towards buying a certain category of products (products tagged/labeled as of a certain category type) – observation: a product has more than one “tag” meaning it has category and one or more subcategories. The real value score is in range 0.00 up to 1.00 where a value of 0 means that the client has never bought that particular category of products, a value of 1.00 means a potential client that buys only that particular category of products and a value around 0.5 means the client buys about half of his/hers products from that particular category of products.

The product segment preference features can be used directly or in a model (ranging from simple statistical models to complex machine learning models) for the following purposes:

- a) Customer behavior segmentation including margin-revenue segmentation
- b) Market-Basket based segmentation or individual market basket inference
- c) Products/Buy recommendation
- d) Churn prediction
- e) Direct use in new product campaigns (for particular product categories), narrowly-targeted campaigns or other kind of campaigns

3 Features derived from Machine Learning models for client price range (level) preference

(obs.: all 15 proposed features have descriptive names instead of short cryptic codes due to the need for high self-explain ability)

Name	Description	Notes
PR01_BABY	Price range score for BABY product category	Real value score [0.00-1.00]
PR02_COSMETICE	Price range score for COSMETICE product category	Real value score [0.00-1.00]
PR03_DERMOCOSM	Price range score for DERMOCOSMETICE product category	Real value score [0.00-1.00]
PR04_PHARMA	Price range score for PHARMA product category	Real value score [0.00-1.00]
PR05_SPECIALE	Price range score for SPECIALE product category	Real value score [0.00-1.00]
PR06_OTC	Price range score for OTC product category	Real value score [0.00-1.00]
PR07_MEDICAT_ALTERN	Price range score for MEDICATIE_ALTERNATIVA product category	Real value score [0.00-1.00]
PR08_BARBATI	Price range score for BARBATI product category	Real value score [0.00-1.00]
PR09_FEMEI	Price range score for FEMEI product category	Real value score [0.00-1.00]
PR11_COPII	Price range score for COPII product category	Real value score [0.00-1.00]
PR12_FRUMUSETI	Price range score for FRUMUSETI product category	Real value score [0.00-1.00]
PR13_HOMEOPATIE	Price range score for HOMEOPATIE product category	Real value score [0.00-1.00]
PR14_SANATATE	Price range score for SANATATE product category	Real value score [0.00-1.00]

PR15_ ANALGEZICE	Price range score for N02_ANALGEZICE product category	Real value score [0.00-1.00]
------------------	---	------------------------------

The client price range features (PR_<CATEG>) define the actual disposition of a client to buy cheap or expensive products from a certain category of products. For example a value of 0.1 for PR09_FEME1 would mean that particular client buys very cheap women-related products while a value of 0.9 of PR09_FEME1 would mean that particular client buys some of the most expensive products in that category (women-related products). This particular feature is applicable in various features but has great potential as a predictor in the churn prediction model. The range of models that can use the client price range features contains the following:

- a) Behavior bases customer segmentation
- b) Client lifecycle analysis and inference
- c) Product/buy recommendation
- d) Churn prediction
- e) Value related segmentation models
- f) Targeted campaigns