

Index

- African minority languages, 134, 153, 166, 183, 191
 Akaike's Information Criterion, 184
 analytical language, 188
 area under the curve, 61, 67, 81, 85, 86, 90, 93, 95,
 111, 179
 AUC. *See* area under the curve
 Automated Readability Index, 2
 automatic classifiers, 184
- Bangla, 20
 bell curve, 184
 binary classifiers, 61, 70, 185
 limitations, 73
 binary logistic classifiers, 66
 BLR. *See* logistic regression, binary
 Britannica corpus, 7
- CCD. *See* China Centre of Disease Control
 CEFR corpus, 17
 chemical names, 153
 China Centre of Disease Control, 35, 169, 193
 Chinese, 13, 169, 193
 Chinese health information complexity
 Index, 179
 Chinese Readability Index Explorer, 171
 classification accuracy, 95
 climate change, 101
 Cloze test, 3, 184
 Coease, 14
 cognitive features, 12
 cognitive style, 101
 coherence, 184
 Coh-Metrix, 13
 Coleman-Liau, 190
 colloquial words, 56
 communication style, 92, 188
 complex words, 58
 computer code, 6
 confidence interval, 98
 confusion matrix, 68
 content words, 174, 178
- continuous variables, 66, 69
 corpus, 60, 62, 92, 184, 185
 feature analysis, 77–79
 feature annotation, 135, 187
 cross-validation, 80, 187
- Dale-Chall readability formula, 2, 5
 Danielson-Bryan 1, 190
 decision making, 116
 deep learning, 9, 184
 Diagnostic Test Calculator, 126
 difficult words, 172, 176
 disaster information, 57, 104, 108, 116, 121, 124,
 125, 132
 Dunn test, 39
- educational resources, 55
 effect size, 90, 92, 188
 effective communication, 55
 emotional interaction, 101
 English, 6, 8, 56, 105, 132
 English-speaking countries, 57
 environmental health information, 57, 61, 70, 75,
 92, 101, 134, 184, 186, 189, 191
 environmental texts, 183
 erroneous translations, 141
 evaluation, 68, 95, 98
 eye-tracking, 6
- Fagan's nomogram, 126, 128, 129
 false negative, 61, 68, 103, 117, 123, 124, 129, 141, 192
 false positive, 61, 68, 103, 123
 familiar expressions, 152, 153
 Fantizi, 171
 feature annotation, 105, 189
 feature normalisation, 81, 86, 88, 93, 95
 feature optimisation, 79, 85, 86, 102
 feature reduction, 187
 Flesch Reading Ease, 2, 4, 5, 8, 13, 14, 15, 18, 20, 21,
 55, 56–58, 66, 67, 69, 70, 72, 73, 110, 126, 127,
 129, 130, 185, 186

Index

227

- Flesch-Kincaid Grade Level, 2, 190
 FORECAST, 3, 4, 66, 67, 68, 69, 70, 73, 185, 186
 forest plot, 97
 formal expressions, 149, 150, 156
 formal words, 101, 137, 153
 French, 8, 12, 18–19
 frequent words, 137
- general public, 186
 Georgia State University, 77
 German, 19
 Google Translate, 156, 183, 192
 Grade Point Level, 183
Guardian newspaper, 7
 Gulpease, 13
 Gunning Fog Index, 2, 57, 58, 59, 66, 67, 68, 71, 73, 185, 186
- health authorities, 60
 health education, 4, 183
 health information, 55, 75, 166
 health professionals, 56
 health translation, 165
 Hindi, 20
 human emotions, 77, 187
 human translators, 129
 hypothesis testing, 184
- Indian languages, 20
 individual differences, 76, 186
 informal expressions, 149
 information load, 174, 177, 184
 interpretability, 79, 85, 90, 93, 95, 102, 186, 189
 Italian, 13–15
- Jaeho Lee, 108
 Japanese, 6, 105, 125, 132, 183, 188, 190
 translations, 108
 Japanese Text Readability Measurement Scale, 108, 109, 121, 125, 129, 131
 Japanese translations, 117
 jargon, 56
 Jiantizi, 172
 joint optimisation, 79
 JTRMS. *See* Japanese Text Readability Measurement Scale
- Kango, 131, 191
 Korp corpus, 17
 Kruskal-Wallis test, 35, 38
- La Repubblica* newspaper, 15
 Lancaster University, 105
 language models, 184
 language style, 92
- Läsbarhets index, 17
 LäsBarT corpus, 16
 Latent Dirichlet Allocation, 63, 186
 LDA. *See* Latent Dirichlet Allocation
 legal language, 5
 less-resourced languages, 165
 level of education, 54
 Lexical categories, 77
 lexical complexity, 101
 Lexical difficulty, 172
 lexico-grammatical features, 78
 likelihood ratio, 115, 116, 121, 122, 123, 124, 129, 141, 190
 linguistic features, 129
 Linsear Write Formula, 21
 log odds, 97
 Logistic Regression, 61, 75, 81, 86, 103, 108, 140, 141, 183, 184, 187, 189, 191
 Binary, 61, 62
 Multinomial, 61
 log-odds, 97
- machine learning, 6, 75, 95, 102, 109, 114, 117, 125, 133, 134, 165, 168, 183, 184, 186, 188, 191
 machine translation, 156, 191, 192
 Mann-Whitney test, 176, 177, 184, 187
 Markov model, 7
 medical abbreviations, 56
 Merck Manual of Diagnosis, 35, 193
 Microsoft Bing Translate, 156, 183, 192
 minority languages, 121, 134
 mistranslations, 123, 151, 153, 156, 165
 MLR. *See* logistic regression, multinomial
 model comparison, 81
 model performance, 86
 model selection, 121
 MSD. *See* Merck Manual of Diagnosis
 multicollinearity, 79
 multiple linear regression, 4, 184
 multiple logistic regression, 95, 186, 188
 multi-syllable words, 137, 149, 152, 154, 156
 multivariable logistic regression, 98
- narrative style, 153, 156
 National Taiwan Normal University, 171
 natural disasters, 188
 NDRISC report, 104
 neural network, 10, 184
 Newsela Corpus, 7
 non-native speakers of English, 56
 non-readable translations, 140
 number of proper nouns, 184
 number of syllables per word, 56
 number of unique monosyllables, 184
 number of words, 184

- odds, 117, 129
- odds ratio, 97, 98, 188
- OneStopEnglish corpus, 7
- online health resources, 58
- optimised model, 86
- Ordsvariationsindex, 15
- orthographic complexity, 174
- Orthographic Levenshtein Distance, 12
- outcome variable, 184
- out-of-dictionary words. *See* out-of-vocabulary words
- Out-of-vocabulary words, 78
- overfitting, 81

- paediatric otolaryngology, 58
- paraphrasing, 154, 156
- parents and caregivers of young children, 57, 75, 186
- parsimony, 85, 95
- Pearson product moment correlation coefficient, 12, 184
- people-centred approach, 60
- percentage of complex words, 150
- percentage of proper nouns, 135
- percentage of unique words, 137, 152
- perplexity, 8
- personal reference, 101
- Philippines, 57
- polysemous words, 172, 177
- polysyllabic words, 59
- Portuguese, 13
- posterior odds, 115, 123, 126
- posterior probability, 115, 117, 125, 126, 129, 190
- precision, 95, 121
- prediction accuracy, 93
- predictive assessment, 122, 125, 134, 141, 149
- predictive modelling, 114, 115, 116
- predictive tools, 124
- predictor variable, 61, 97, 98, 100, 184
- pre-editing, 134, 151, 165, 191
- principles of health communication, 54
- prior odds, 122
- prior probability, 115, 117, 126, 129, 190
- probability threshold, 103, 113, 114, 115, 116, 117, 122, 123, 124, 125, 129, 141, 156, 165, 191, 192
- psycholinguistic measure, 14
- p-value, 90, 184

- readability, 117, 121, 124, 125, 129, 183, 185
- Readability Studio, 78, 105, 135, 187, 189, 191
- readability tools, 54, 66, 76
- readable translations, 140
- READ-IT, 14
- Receiver Operating Characteristic, 67, 68, 111, 116, 140, 185

- Recurrent Neural Network, 11
- recursive feature elimination, 85, 88, 90, 187
- ROC. *See* receiver operating characteristic

- school grade, 66, 69
- screening, 124
- SEANCE. *See* Sentiment Analysis and Social Cognition Engine
- semantic analysis, 77
- semantic code, 187
- Semantic complexity, 172
- semantic difficulty, 59
- semantic features, 86, 88, 90
- sensitivity, 61, 68, 69, 70, 85, 86, 88, 90, 93, 103, 114, 116, 117, 123, 129, 141, 165, 185
- sentence length, 56, 58, 131
- Sentiment analysis, 77
- Sentiment Analysis and Social Cognition Engine, 77, 85, 87–90, 187
- sentiment features, 85, 88
- Sequential Minimal Optimisation, 7
- short words, 56
- shorter sentences, 154
- significance level, 98
- Simple Measure of Gobbledygook, 3, 59–60, 66, 68, 137, 149, 155, 185, 186
- simple regression, 184
- skin conditions, 65
- Slovenian, 7, 13
- SMOG. *See* Simple Measure of Gobbledygook
- Somali, 134, 141, 155
- specific readerships, 55, 60, 185
- specificity, 61, 68, 69, 70, 85, 86, 88, 90, 93, 103, 114, 116, 117, 123, 129, 141, 165, 185
- standard deviation, 97
- standardised coefficient beta, 97
- statistical feature selection, 187
- strokes per character, 173, 177, 193
- structural features, 78, 86, 88, 90
- structural simplicity, 189
- Support Vector Machine, 17, 109, 117, 121, 140, 183, 184, 187, 189
- SVM. *See* Support Vector Machine
- Swahili, 134, 141, 155
- Swedish, 15–17
- syllables, 14
- syllables per word, 184
- syntactic difficulty, 59
- syntactic parsing, 184

- t test, 95, 179
- target audiences, 55, 186
- target language, 124
- technical jargon, 156
- text revision, 126, 149, 150, 151, 153, 155

Index

229

- text simplification, 126
- thematic topics, 63
- thematic variation, 65, 186
- threshold values, 69
- Tigrinya, 134, 141, 155
- timeliness, 72
- topic model, 63, 64, 186
- training dataset, 129
- translatability, 116, 132, 135, 141, 149, 153, 154, 156, 165
- translatable sentence, 129
- translation, 125
- translation flow, 134
- translation quality, 104, 123, 124, 134, 165, 190
- translation studies, 167
- true negative, 68, 117
- true positive, 68
- TTR. *See* Type-Token Ratio
- Type I error, 68, 70–72, 73
- Type II error, 70, 72, 73
- Type-Token Ratio, 184
- UCREL Semantic Analysis System, 77, 86–90, 105, 135, 187, 189, 191
- understandability, 53, 54
- US Navy, 3
- USAS. *See* UCREL Semantic Analysis System
- Wago, 132, 191
- Waseda University, 108
- web page summaries, 5
- WeeBit corpus, 7
- Weeklyreader corpus, 6
- WHO. *See* World Health Organization
- WOCADI parser, 19
- word frequency list, 172
- World Health Organization, 35, 54, 72, 101, 169, 185, 193
- zero-importance feature elimination, 79, 85
- Zipf's Law, 55, 56
- z-score, 187
- Zulu, 135, 141