

Оглавление

Предисловие	3
Введение	4
1. Функциональные возможности платформы PolyAnalyst	10
1.1. Архитектура платформы PolyAnalyst	10
1.2. Масштабируемость и производительность	12
1.3. Форматы и источники анализируемых данных. Коннекторы к данным	13
1.4. Возможности анализа текстовых данных. Машинное обучение и язык XPDЛ для описания и извлечения лингвистических паттер- нов	14
1.5. Инструменты предобработки и очистки данных (ETL)	16
1.6. Инструменты анализа числовых данных. Data Mining, машин- ное обучение (ML), подключение внешних библиотек и кода на языках Python и R	17
1.7. Инструменты визуализации для создания аналитических и свод- ных панелей (блок BI)	21
1.8. Экспорт результатов и отчетов	24
1.9. Интерфейс PolyAnalyst API для интеграции с ИТ-системами и привязки к бизнес-процессам заказчика	25
2. Установка PolyAnalyst	26
3. Начало работы с PolyAnalyst	31
3.1. Запуск сервера PolyAnalyst	31
3.2. Вход в систему	35
3.3. Стартовая страница	35
3.4. Управление персональными данными в профиле пользователя ..	36
3.5. Аналитический клиент	37
4. Импорт данных	39
4.1. Виды узлов-источников данных	39
4.2. Настройки узлов импорта данных	41
Узел <i>Файлы CSV</i>	41
Узел <i>Файлы</i>	48
Узел <i>Изображения</i>	50
Узел <i>XML</i>	51
Узел <i>Microsoft Excel</i>	52
Узел <i>ODBC</i>	53
Узел <i>OLEDB</i>	57
Узел <i>Интернет</i>	57

Узел <i>RSS</i>	60
Узел <i>SharePoint</i>	60
4.3. Импорт данных в случае установки серверной части PolyAnalyst на удаленной машине.....	62
5. Экспорт данных	63
5.1. Виды узлов экспорта данных.....	63
5.2. Настройки узлов экспорта данных.....	64
Узел <i>Экспорт в файл (CSV)</i>	64
Узел <i>Экспорт в MS Word</i>	69
Узел <i>Экспорт в папку</i>	70
Узел <i>Экспорт в словарь</i>	71
Узел <i>Отправка писем</i>	73
Узел <i>Экспорт в изображение</i>	75
Узел <i>Создание публикации</i>	76
Узел <i>Экспорт в ODBC</i>	77
6. ETL. Манипуляции с данными	79
6.1. Виды узлов операций с колонками.....	79
6.2. Настройки узлов операций с колонками.....	80
Узел <i>Анонимайзер</i>	80
Узел <i>Дискретизация</i>	81
Узел <i>Нормализация</i>	82
Узел <i>Консолидация категорий</i>	83
Узел <i>Извлечение терминов</i>	85
Узел <i>Фильтрация колонок</i>	87
Узел <i>Модификация колонок</i>	87
Узел <i>Замена категорий</i>	89
Узел <i>Замена терминов</i>	91
Узел <i>Оптическое распознавание символов</i>	92
Узел <i>Производные колонки</i>	93
Узел <i>Применение моделей</i>	94
Узел <i>Объединение моделей</i>	94
6.3. Виды узлов операций со строками.....	95
6.4. Настройки узлов операций со строками.....	96
Узел <i>Смещенная выборка</i>	96
Узел <i>Дополнение</i>	96
Узел <i>Обогащение данных</i>	97
Узел <i>Уникальные записи</i>	98
Узел <i>Фильтрация строк</i>	99
Узел <i>Выборка</i>	100
Узел <i>Суммы и ранги</i>	101
6.5. Виды узлов операций с таблицами.....	102
6.6. Настройки узлов операций с таблицами.....	103
Узел <i>ABC XYZ анализ</i>	103
Узел <i>Python</i>	104

Узел <i>Агрегирование</i>	105
Узел <i>Конкатенация</i>	107
Узел <i>Объединение</i>	108
Узел <i>Применение таксономии</i>	109
Узел <i>Производная таблица</i>	111
Узел <i>Сравнение</i>	111
Узел <i>Структурная организация</i>	112
Узел <i>Транспонирование</i>	112
7. Текстовый анализ	114
7.1. Использование словарей.....	115
7.2. Виды узлов для текстового анализа.....	120
7.3. Настройки узлов текстового анализа.....	122
Узел <i>Автоматическая таксономия</i>	122
Узел <i>Анализ тональности</i>	124
Узел <i>Анализ трендов</i>	125
Узел <i>Анонимизация сущностей</i>	126
Узел <i>Восстановление сущностей</i>	127
Узел <i>Извлечение ключевых слов</i>	128
Узел <i>Извлечение сущностей</i>	131
Узел <i>Извлечение медицинских сущностей</i>	132
Узел <i>Извлечение сущностей с обучением</i>	132
Узел <i>Извлечение фактов</i>	133
Узел <i>Индекс</i>	134
Узел <i>Классификация на основе таксономии</i>	135
Узел <i>Классификация текстов</i>	136
Узел <i>Кластеризация текстов</i>	139
Узел <i>Объединение таксономий</i>	141
Узел <i>Определение языка</i>	142
Узел <i>Перевод текстов</i>	143
Узел <i>Поисковый запрос</i>	144
Узел <i>Проверка грамматики</i>	145
Узел <i>Проверка орфографии</i>	146
Узел <i>Разметка текста</i>	147
Узел <i>Расшифровка сокращений</i>	149
Узел <i>Резюме</i>	151
Узел <i>Связь терминов</i>	151
Узел <i>Сравнение текстов</i>	152
Узел <i>Схожесть текстов</i>	153
Узел <i>Таксономия</i>	153
Узел <i>Тематика текстов</i>	155
Узел <i>Тональность текстов</i>	155
Узел <i>Уникальные тексты</i>	156
Узел <i>Удаление фрагментов</i>	158
8. Анализ данных и алгоритмы машинного обучения	159

8.1. Алгоритмы машинного обучения и другие узлы PolyAnalyst, применяемые для анализа данных	159
8.2. Настройки узлов анализа данных и машинного обучения	164
Узел <i>ARIMA</i>	165
Узел <i>SVM-классификатор</i>	166
Узел <i>Адаптивный бустинг</i>	167
Узел <i>Многомерные адаптивные регрессионные сплайны</i>	167
Узел <i>Сверточная нейронная сеть</i>	168
Узел <i>Факторный анализ</i>	169
Узел <i>Анализ покупательских корзин</i>	170
Узел <i>Анализ связей</i>	171
Узел <i>Анализ социальных сетей</i>	171
Узел <i>Нейронная сеть</i>	172
Узел <i>Сеть Козонена</i>	173
Узел <i>Анализ статистических распределений</i>	174
Узел <i>Аудит данных</i>	175
Узел <i>Байесовская классификация</i>	175
Узел <i>Нечеткое объединение</i>	176
Узел <i>Ближайшие соседи</i>	176
Узел <i>Случайный лес решений</i>	177
Узел <i>Кластеризация на основе расстояния</i>	177
Узел <i>Таблица решений</i>	178
Узел <i>Дерево решений</i>	179
Узел <i>Нечеткое совпадение</i>	179
Узел <i>Кластеризация по методу K-средних</i>	180
Узел <i>Корреляционный анализ</i>	180
Узел <i>Линейная классификация</i>	181
Узел <i>Нормализация сущностей</i>	182
Узел <i>Тестирование модели</i>	182
Узел <i>Логистическая регрессия</i>	183
Узел <i>Максимизация ожидания</i>	183
Узел <i>Транзакционный анализ покупательских корзин</i>	184
Узел <i>Прогнозирование временных рядов</i>	184
9. Инструменты визуализации в платформе PolyAnalyst	186
9.1. Виды узлов визуализации данных	186
9.2. Настройки узлов визуализации данных	188
Узел <i>OLAP-Таблица</i>	188
Узел <i>Агрегированный график</i>	190
Узел <i>Граф</i>	190
Узел <i>Диаграмма рассеяния</i>	191
Узел <i>Диаграмма связей</i>	193
Узел <i>Иерархическая диаграмма</i>	194
Узел <i>Карта</i>	195
Узел <i>Коробчатая диаграмма</i>	197

Узел <i>Линейный график</i>	199
Узел <i>Матрица диаграмм рассеяния</i>	200
Узел <i>Многомерная матрица</i>	201
Узел <i>Облако тегов</i>	201
Узел <i>Объект GIS</i>	202
Узел <i>Параллельные координаты</i>	203
Узел <i>Пузырьковая диаграмма</i>	204
10. Разработка интерактивных графических аналитических панелей (дашбордов)	205
10.1. PolyAnalyst как BI-система. Зачем нужен «генеральский слайд»	205
10.2. Примеры дашбордов, построенных в PolyAnalyst	206
10.3. Процесс создания аналитического отчета	208
10.4. Работа в редакторе отчетов PolyAnalyst. Как создать шедевр ...	211
10.5. Добавление в отчеты внешних виджетов	219
Заключение	222
Литература	223