Всё сильно зависит от сроков, в которые это нужно разработать, но если считать, что я не особо ограничен во времени, то:

* Для начала создам упрощенный алгоритм, который будет искать слова по корням / основам нецензурных слов. В базе перечислю все слова, которые мне удастся найти в словарях матерных слов и вообще в интернете, а так же их корни / основы.
* Сделаю кнопку «пожаловаться на нецензурную лексику» что бы пользователи обращали мое внимание на те фразы, которые пропустил фильтр, а так же жалоба на неправильное определение фразы, как нецензурную (в случае, если алгоритм сработал).
* Сделаю отдельную web мини-игру «Перематери Петровича 18+», в которой нужно будет удивлять Петровича новыми фразами, которые он еще на знает. В ответ Петрович будет удивлять пользователя. Все новые фразы попадают в базу, и потом проходят пост-модерацию. На постмодерацию будут попадать фразу, по которым Google или Яндекс возвращают более n результатов, остальные игнорироваться. В игру будет присутствовать лидерборд.

Инструменты: Python, MySQL, Django

Часть 2.

Когда заработает первая часть, я займусь второй частью с уклоном в ML и NLP. Начну с того, что попробую обучить модель для классификации слов по признаку цензурные / нецензурные. База с целевыми признаками должна будет пополняться из первой части мой разработки. Тут я попробую обучать NLP библиотеки, а кодить буду на Python.

В зависимости от успеха / неуспеха NLP моделей попробую обучить нейросеть на определение нецензурных слов и выражений.

Основные этапы:

* Выяснение требований заказчика, Уточнение ТЗ, анализ и оценка этих требований;
* Построение теоретической модели. Описание структура базы и основной логики работы сервиса;
* Сбор данных в интернете (словарь нецензурных слов и выражений) и доведение их до приемлемого состояния (формирование корней / основ / исключений);
* Кодирование;
* Тестирование, поиск и устранение багов.
* Внедрение и анализ работы на продакшне, доработка проекта.\

Лучшим методом тестирование (если у меня нет команды тестировщиков) – считаю геймефикацию и предоставление возможности неограниченному кол-ву пользователей протестировать продукт в направлении «переиграть систему».

Оценить качество инструмента можно по парсингу ЖЖ Артемия Лебедева и прогонки через систему большого кол-ва фраз из не цензурируемых источников (форумов, соц. сетей и.т.д.)