**Отчет по определению факторов влияющих на тип и качество вина**

1. **Введение**  
   В соревновании [kaggle](https://www.kaggle.com/rajyellow46/wine-quality) требуется предсказать тип вина исходя из 13 полей и определить наиболее значимые химические вещества необходимые для формирования качественного вина на производстве.
2. **Взаимосвязь факторов**

В ходе исследования была определена взаимосвязь факторов, представленная на рисунке ниже:

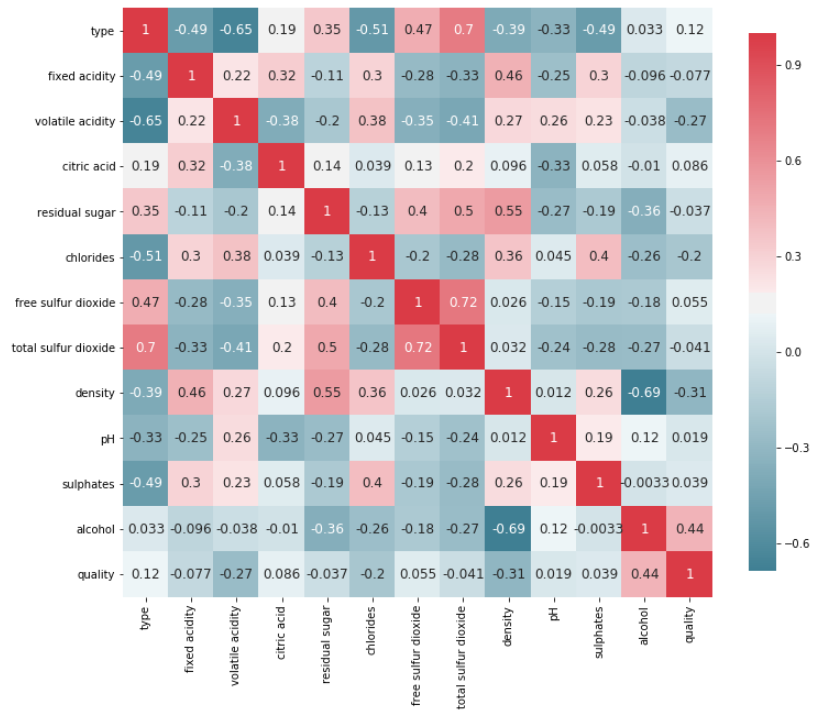


Рисунок 1 – Коррелограмма факторов

Из коррелограммы видно, что наибольшее влияние на качество вина оказывает % алкоголя в вине. Соответвенно важно придерживаться оптимальной технологии. Этапы технологического процесса следует исследовать дополнительно.

**3. Важность факторов**

Важность факторов производилась на основе алгоритма «дерева решений». Для данной выборки получили следующие результаты:

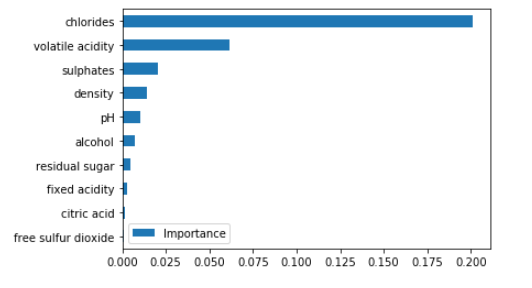


Рисунок 2 – Важность факторов при прогнозировании типа вина

1. **Выбор модели для предсказания типа вина**

Мы были очень ограничены во времени проведения исследования, поэтому использовали Логистическую регрессию, которая показала отличные результаты как на тестовой так и на валидационной выборке (score = 98% на обоих выборках).

1. **Обсуждение улучшения модели**

В случае необходимости дальнейшего улучшения качества классификации следует использовать модели большего порядка (staking, boosting, bagging) и проводить подбор гиперпараметров моделей через GridSeachCV.

1. **Дальнейшее развитие модели**

Текущих показателей модели достаточно чтобы внедрить ее в процесс производства вина на начальном этапе. Планируем внедрить как есть, проверить и вернуться к улучшению показателей модели после внедрения остальных моделей из программы развития Data science в компании