

Mineração de Dados

Prof. Rodrigo Tripodi Calumby - rtcalumby@uefs.br

Desafio Experimental 2 - Descobrimdo Padrões, Regras e Recomendações

Tema: Futebol Mundial 2018

Contexto do Desafio

O futebol é sem dúvida uma grande paixão mundial. Mais do que isso, é um grande negócio, movimentando fortunas ao redor do mundo. Neste contexto, tanto a paixão quanto o dinheiro estão associados ao desempenho dos jogadores e seus times. Como uma grande tendência, as diversas categorias de esporte exploram cada vez mais os dados disponíveis para extrair conhecimento decisivo em termos tanto de performance esportiva, quanto retorno de investimentos. Desse modo, considerando dados coletados sobre um grande número de jogadores, este desafio visa o desenvolvimento de soluções de mineração de dados que permitam a descoberta de padrões não triviais para o entendimento dos diferentes cenários e a tomada boas decisões.

Objetivo Geral

A equipe deverá apresentar 3 soluções de mineração de dados ou 1 sistema interativo baseado em:

1. Aplicação de técnicas de agrupamento. Exemplos:
 - a) Análise de grupos: características, padrões frequentes, estatísticas (momentos, correlações, etc).
 - b) Descoberta de jogadores fora dos padrões gerais.
2. Descoberta de padrões frequentes e regras de associação fortes. Exemplos:
 - a) Quais padrões existem em diferentes contextos (nacionalidades, posição, campeonatos, entre outros).
 - b) Quais regras de associação podem ser inferidas em cada cenário?
 - c) Análise de cruzamento de regras entre cenários.
 - d) Existem padrões e regras fortes de modo geral?
3. Recomendação de jogadores. Exemplos:
 - a) Quais os jogadores com características parecidas? (contração para substituir jogador vendido; redução de custos com jogador parecido, mas com salário menor; substituição temporária de jogador lesionado).
 - b) Dado um cenário, existem jogadores para a mesma função (ataque, defesa, esquerda, direita, etc.) e que tenham características complementares?

Objetivos Específicos

A equipe deve realizar o pré-processamento necessário dos dados para torná-los adequados ao processo de mineração (limpeza, tratamento de dados ausentes, padronização, conversão e seleção de atributos, etc.). Para as soluções propostas, as equipes devem:

1. Definir qual(is) algoritmos serão utilizados/desenvolvidos para cada solução;
2. Definir a metodologia experimental de execução/validação das propostas (configurações dos algoritmos, protocolos experimentais, medidas de eficácia, etc.);
3. Definir os modelos a serem utilizados para visualização de resultados.

Objetivo Extra (opcional)

Como atividade extra, poderá ser realizada a integração de novos dados que possam contribuir para o aumento da eficácia, diversidade ou informatividade das soluções propostas. Como sugestão, podem ser coletados mais dados sobre os jogadores, por exemplo, se foi convocado para a copa 2018 ou não, número da camisa, número de gols na temporada, etc. Novos dados podem também ser derivados dos dados já disponíveis. Haverá pontuação bônus para este objetivo extra de acordo com a qualidade das contribuições apresentadas.

Produtos

A equipe deve entregar o pré-projeto (1 página) na aula do dia 26/08. A equipe deverá entregar todo os códigos, *scripts*, *workflows*, dados, etc. que forem utilizados/desenvolvidos no projeto. Além disso, a equipe deverá apresentar um relatório, incluindo pelo menos:

- Introdução
- Metodologia
- Resultados e Discussão
- Conclusões
- Referências

Os materiais utilizados/desenvolvidos devem ser entregues por e-mail em arquivo compactado. Este material deve estar organizado em subpastas e com nomes significativos, por exemplo, “../dados/instagram.csv”, “../scripts/normalizaçãoAtributoX.sh”, “../workflows/experimento1.ows”, etc. O relatório deve ser entregue no mesmo e-mail em formato PDF (nome do arquivo: “relatórioEquipeX.pdf”). O relatório deve ser produzido em formato de artigo científico, seguindo o [modelo de formatação IEEE Transactions](#). Utilize preferencialmente o *template* LaTeX. O relatório deve ter no máximo 6 páginas (incluindo imagens e referências). O uso de páginas adicionais podem ser solicitadas ao professor, com a devida justificativa.

Os produtos devem ser enviados para o endereço `rtcalumby(arroba)ecomp(ponto)uefs(ponto)br` **até as 08h do dia 16/07/18**. O assunto do e-mail deve ser: [Mineração 2018.1] Entrega Desafio Prático 2 - Equipe X.

Critérios Gerais de Avaliação

- Atendimento ao prazo de entrega estabelecido.
- Completude do material entregue.
- Atendimento aos objetivos, requisitos e orientações desta proposta.
- Adequação do protocolo experimental aplicado aos dados utilizados.
- Amplitude dos experimentos realizados.
- Rigor científico dos experimentos realizados e resultados apresentados.
- Qualidade da discussão dos resultados e conclusões.
- Qualidade da apresentação visual do relatório.
- Atendimento à formatação indicada.
- Qualidade do texto do relatório:
 - Clareza e objetividade.
 - Atendimento às normas de ortografia e gramática.
 - Adequação das citações.
 - Organização das informações.

Observações

- No relatório, só deve ser feita descrição de métodos não triviais/comuns e que tenham sido utilizados no projeto. Não deve ser feita fundamentação teórica daquilo que já foi visto em sala de aula.
- A critério do professor, alunos poderão ser convocados para responder a questionamentos sobre o trabalho desenvolvido.
- Fraudes serão penalizadas com nota 0.0 (zero) na atividade.
- Entregas com atraso de até 24h serão penalizadas em 2 (dois) pontos. Após 24h, entregas não serão mais aceitas, a menos que seja comprovado motivo de força maior que comprometa toda a equipe.

Anexo I - Dicionário de Dados

Dataset

17994 jogadores, 185 atributos.

Informações básicas

ID, name, full_name, club, club_logo, special, league, flag, nationality, photo.

Características físicas

Numéricos: age, height_cm, weight_kg.

Catégoricos: birth_date, body_type.

Booleano: real_face

Dados financeiros (numéricos):

eur_value, eur_wage, eur_release_clause

Características gerais do jogador

Numéricos: overall, potential, pac, sho, pas, dri, def, phy, international_reputation, skill_moves, weak_foot.

Catégoricos: work_rate_att, work_rate_def, preferred_foot.

Habilidades do jogador (numéricos)

crossing, finishing, heading_accuracy, short_passing, volleys, dribbling, curve, free_kick_accuracy, long_passing, ball_control, acceleration, sprint_speed, agility, reactions, balance, shot_power, jumping, stamina, strength, long_shots, aggression, interceptions, positioning, vision, penalties, composure, marking, standing_tackle, sliding_tackle, gk_diving, gk_handling, gk_kicking, gk_positioning, gk_reflexes.

Pontecial do jogador por posição (numéricos)

rs, rw, rf, ram, rcm, rm, rdm, rcb, rb, rwb, st, lw, cf, cam, cm, lm, cdm, cb, lb, lwb, ls, lf, lam, lcm, ldm, lcb, gk.

Características adicionais do jogador (booleanos)

1_on_1_rush_trait, acrobatic_clearance_trait, argues_with_officials_trait, avoids_using_weaker_foot_trait, backs_into_player_trait, bicycle_kicks_trait, cautious_with_crosses_trait, chip_shot_trait, chipped_penalty_trait, comes_for_crosses_trait, corner_specialist_trait, diver_trait, dives_into_tackles_trait, diving_header_trait, driven_pass_trait, early_crosser_trait, fan's_favourite_trait, fancy_flicks_trait, finesse_shot_trait, flair_trait, flair_passes_trait, gk_flat_kick_trait, gk_long_throw_trait, gk_up_for_corners_trait, giant_throw_in_trait, inflexible_trait, injury_free_trait, injury_prone_trait, leadership_trait, long_passer_trait, long_shot_taker_trait, long_throw_in_trait, one_club_player_trait, outside_foot_shot_trait, playmaker_trait, power_free_kick_trait, power_header_trait, puncher_trait, rushes_out_of_goal_trait, saves_with_feet_trait, second_wind_trait, selfish_trait, skilled_dribbling_trait, stutter_penalty_trait, swerve_pass_trait, takes_finesse_free_kicks_trait, target_forward_trait, team_player_trait, technical_dribbler_trait, tries_to_beat_defensive_line_trait, poacher_speciality, speedster_speciality, aerial_threat_speciality, dribbler_speciality, playmaker_speciality, engine_speciality, distance_shooter_speciality, crosser_speciality, free_kick_specialist_speciality, tackling_speciality, tactician_speciality, acrobat_speciality, strength_speciality, clinical_finisher_speciality.

Posições de preferência do jogador (booleanos)

prefers_rs, prefers_rw, prefers_rf, prefers_ram, prefers_rcm, prefers_rm, prefers_rdm, prefers_rcb, prefers_rb, prefers_rwb, prefers_st, prefers_lw, prefers_cf, prefers_cam, prefers_cm, prefers_lm, prefers_cdm, prefers_cb, prefers_lb, prefers_lwb, prefers_ls, prefers_lf, prefers_lam, prefers_lcm, prefers_ldm, prefers_lcb, prefers_gk.