

## Review Makalah

Judul: *Pattern Recognition Techniques for Provenance Classification of Archeological Ceramics using Ultrasonic*

Author: Addisson Salazar, Gonzalo Safont, Luis Vergara, Enrique Vidal

## Masalah

Makalah ini mengulas tentang penggunaan pengenalan pola untuk klasifikasi keramik-keramik kuno yang ditemukan di situs-situs sejarah berdasarkan karakteristik keramik. Pada makalah ini karakteristik keramik dideteksi dengan gelombang ultrasonik. Fitur-fitur gelombang ultrasonik yang dijadikan representasi dari karakteristik keramik adalah *Total signal attenuation*, *Propagation velocity*, *Signal power*, *Attenuation curve initial value*, *Principal frequency*, *Principal frequency amplitude*, *Centroid frequency*, *Instantaneous centroid frequency*, *Time-reversibility*, dan *Third order autocovariance*.

Secara umum masalah klasifikasi pada makalah ini dapat dituliskan sebagai berikut. Misalkan  $K$  adalah kumpulan fitur-fitur gelombang ultrasonik,  $X$  adalah kumpulan *instance* karakteristik keramik dan  $Y$  adalah kumpulan jenis-jenis keramik. Label-label yang digunakan adalah  $Y_1 = \{\textit{terra sigillata}, \textit{non terra sigillata}\}$ ,  $Y_2 = \{\textit{paterna}, \textit{les jovaes}\}$ , dan  $Y_3 = \{\textit{terra sigillata Aretina}, \textit{terra sigillata Sud Gaul}\}$ . Untuk mengklasifikasi keramik ke dalam  $Y_1$ , dibutuhkan  $X_1$  yang berisi semua *instance* fitur  $K_1 = K$ . Pada klasifikasi  $Y_2$ , dibutuhkan  $X_2$  yang berisi semua *instance* fitur  $K_2 = K - \{\textit{time reversibility}, \textit{principal frequency}\}$ . Pada klasifikasi  $Y_3$ , dibutuhkan  $X_3$  yang berisi semua *instance* fitur  $K_3 = K - \{\textit{centroid}, \textit{maximum frequencies}\}$ . Dengan demikian untuk mengklasifikasi keramik-keramik tersebut diperlukan fungsi target  $f_i : X_i \rightarrow Y_i$ . Fungsi  $f_i$  ini didekati dengan fungsi hipotesis yang dicari dari himpunan fungsi hipotesis  $H_i = \{h_i | X_i \rightarrow Y_i\}$ . Fungsi hipotesis  $h$  ini dihasilkan dari model yang dikonstruksi dari algoritma pembelajaran.

## Pendekatan Solusi

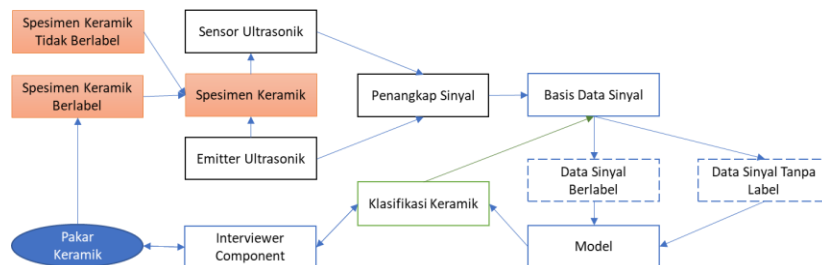
Pendekatan yang digunakan dalam proses pengenalan pola untuk klasifikasi keramik kuno adalah *semi supervised active learning* atau SSAL. Secara umum cara kerja SSAL dapat dituliskan sebagai berikut.

1. Siapkan data sinyal keramik berlabel  $L$  dan data sinyal keramik tanpa label  $U$ .
2. Gunakan  $L$  sebagai data latih model klasifikasi  $C$ .
3. Gunakan  $C$  untuk mengklasifikasi  $U$ .
4. Untuk tiap anggota  $U$ , pilih  $u$  subset dari  $U$  dengan tingkat kepercayaan paling rendah, dan dapatkan label untuk anggota  $u$  dari pakar.
5. Tambahkan  $u$  ke  $L$ .
6. Hapus  $u$  dari  $U$ .
7. Jika  $U$  belum habis, ulangi algoritma ini dari tahap 2.
8. Tampilkan hasil klasifikasi  $L$ .

Pendekatan SSAL ini dinilai tepat karena dapat meminimalisasi kebutuhan peran pakar keramik. Ketika proses klasifikasi berjalan, peran pakar hanya dibutuhkan untuk memberikan input klasifikasi pada kelompok hasil klasifikasi yang paling meragukan. Dengan input ini, sistem dapat belajar sehingga diharapkan dapat memperbaiki klasifikasi pada kelompok hasil dengan tingkat keraguan yang lebih kecil. Namun demikian, untuk mendapatkan meminimalisasi peran pakar, model perlu diuji coba beberapa kali sehingga didapatkan titik optimal antara jumlah data latih dan jumlah input dari pakar.

Metode klasifikasi yang digunakan pada percobaan ini adalah linear discriminant analysis (LDA), random forest (RF), and support vector machine (SVM). Ketiga *classifier* ini adalah one class classifier, yaitu metode yang hanya dapat mengenali satu kelas tertentu.

### Arsitektur sistem



Data yang digunakan untuk membuat model klasifikasi adalah data sinyal ultrasonik yang berasal dari respon material yang ditembak dengan sinyal ultrasonik tersebut.

Data sinyal diperoleh dari dari spesimen keramik. Spesimen keramik dikumpulkan dan beberapa diantaranya diberi label oleh pakar keramik. Kemudian spesimen-spesimen ini ditembak dengan sinyal sinyal ultrasonik dari emitter dan refleksi pantulan sinyal ini ditangkap oleh sensor ultrasonik. Sinyal ini kemudian diambil fitur-fiturnya dan dikumpulkan pada basis data sinyal yang berisi *instance* X. Sinyal yang berasal dari spesimen berlabel diberikan label sesuai dengan label spesimen yang terkait berdasarkan label yang ada pada himpunan label Y. Data sinyal ini kemudian dijadikan data latih untuk membuat model klasifikasi berdasarkan ketiga metode tersebut di atas.

Model dijalankan dalam beberapa iterasi sehingga semua data sinyal mendapatkan label. Pada interaksi pertama, model akan belajar dari data sinyal berlabel dan kemudian mengklasifikasikan data tanpa label. Setelah klasifikasi selesai, model akan memilah klasifikasi dengan tingkat kepercayaan paling rendah dan meminta input dari pakar. Input ini kemudian disimpan dalam kelompok data berlabel. Pada interaksi berikutnya, data berlabel yang “baru” menjadi data latih model untuk proses klasifikasi berikutnya. Interaksi ini berhenti ketika semua data sinyal tanpa label berhasil diklasifikasikan.

### Ulasan

Data mentah yang digunakan pada percobaan ini menggunakan pecahan keramik. Pecahan ini boleh jadi memiliki kualitas material yang berbeda dengan keramik yang masih utuh dengan label yang sama. Keramik dibuat dari material komposit (tanah liat, mineral, dll.). Pada keramik yang utuh, sisi dalam keramik cenderung dalam kondisi yang lebih baik dibandingkan dengan sisi dalam keramik yang pecah. Karena model dibuat tanpa menyertakan data latih dari keramik utuh, dapat diajukan hipotesis bahwa model ini tidak dapat memberikan klasifikasi dengan tingkat kepercayaan yang sama pada klasifikasi keramik utuh. Untuk menangani hal ini, perlu ditambahkan keramik utuh sebagai bahan persiapan data latih.

Penggunaan one class classifier pada percobaan ini dapat dinilai tepat untuk menentukan apakah suatu spesimen keramik memiliki kriteria tertentu atau tidak. Akan tetapi jika spesimen berasal dari lokasi yang merupakan pertemuan dari berbagai jenis budaya, proses klasifikasi harus dilakukan berulang kali. Hal ini dapat diatasi dengan menggunakan algoritma yang menghasilkan model *multi class classifier*. Namun demikian, penggunaan model yang dapat mengenali banyak kelas membutuhkan proses persiapan yang lebih panjang. Hal ini mencakup kebutuhan data latih yang lebih lengkap dan melibatkan pakar yang lebih banyak.